" Thou shalt not work alone "

Damien Besancenot, Kim Van Huynh, Francisco Serranito

“Thou shalt not work alone”: Individual Productivity in Research and Co-authorship in economics

Damien Besancenot, CEPN and University Paris 13, Sorbonne Paris Cité
E-mail: besancenot.damien@univ-paris13.fr

Kim Huynh, LEM and UniversityPanthéon-Assas, Paris 2
E-mail : Kim.Huynh@u-paris2.fr

Francisco Serranito, Univ.Orléans, CNRS, LEO and IRD, LEDa, DIAL UMR 225
E-mail : francisco.serranito@univ-orleans.fr

Abstract : This paper focuses on the properties of the matching process which leads to scientific collaboration. In a first step, it proposes a simple theoretical model to describe the intertemporal choice of researchers facing successive opportunities of co-authoring papers. In a second part, the paper empirically assesses the properties of the model. The main empirical result is that the number and the productivity of a researcher's co-authors reflect the productivity of this researcher. This result is consistent with the assumption that co-authorship is motivated by a willingness to increase both the quality and the quantity of research output. As researchers with a lot of influent publications papers may create links with a large number of influential co-authors, co-authoring with highly productive academics appears as a signaling device of researchers’ quality.

Keywords: Co-authorship, matching, researchers’ strategy.

JEL Classification : A14, C78, C46

Introduction:

In its traditional approach, sociology of science considers research as a solitary activity. Science progresses as a winner takes all race and all the prestige of a breakthrough is granted to the first author who publishes a new result. Co-authorship which limits the prestige of being at the origin of a new concept seems thus unnatural (Stephan 1996). Moreover, co-authorship requires coordination efforts, imposes compromises between authors or may limit the innovative content of the collective work as authors may have different degree of risk aversion (Hudson 1996). Co-authorship may also suffer from free riding, the workload may be unfairly distributed and it is difficult ex ante to identify the true ability of a potential coauthor (Hollis 2001, Fafchamps et al. 2010). Last but not least, the fact that an author writes only co-authored papers is usually seen by his peers as a hint of his/her inability to write a paper alone and is therefore considered as a negative signal of scientific proficiency.

Given the problems raised by scientific collaborations it may seem difficult to understand the monotonic long run increase in co-authorship.

Since the end of the seventies, a vast strand of research emphasizes that coauthoring papers is not the exception but constitutes a new scientific norm (see for instance Beaver and Rosen 1978; Stefaniak 1982; Petry 1988; Zitt et al. 2000, Laband and Tollison 2000, Cardoso et al 2010, Card and DellaVigna 2013, Hamermesh 2013 and 2015)[1]. Generally, the economic literature explains this evolution through the positive effects of scientific collaboration on the quantity and the quality of the research output. Even if empirical evidence is sometimes contradictory, several hints seem to connect co-authorship and improved papers' quality. For instance, Laband (1987) or Johnson (1997) report that citations frequency is significantly higher for co-authored papers compared to single-authored ones. Laband and Tolisson (2000) and Ursprung and Zimmer (2007) document that co-authorship increases acceptance rate by refereed journals. Chung et al. (2009) shows that papers co-authored with a prolific author receive more citations and thus seem to be of a better quality. Co-authoring is also considered as a way to increase the number of papers that a researcher may publish during a given period of time. Durden and Perry (1995) finds that the total number of publications is significantly and positively related to the number of collaborative publications. Hollis (2001) shows that the more co-autorship done in the past, the more prolific an author is likely to be today. Lee and Bozeman (2005) stresses that collaboration is a strong predictor of the total number of a researcher's publications.[2]

As co-authorship is likely to play a central role in the production of knowledge, it is important to understand the factors that favor or hinder collaboration. However, if the motivations of team formation in scientific activities, and more specifically in the writing of papers, are now well identified (see Bruno 2014 for a recent survey), little has been done yet to understand the specific characteristics which lead researchers to choose each other in the building of a new team.

[1] See also Wuchty et al 2007 for a description of the way how in sociology of science, focus shifted from individual researcher to teamwork.

[2]Note that the available empirical evidence on the benefits of co-authorship on individual research productivity is mixed. The results put forward by Lee and Bozeman (2005) or Hollis (2001) disappear once the scientific production is weighted by the number of co-authors. Chung et al. (2009) doesn't disclose any effect of co-authorship on quality when papers are written with colleagues at the same institution (see Bidault and Hilldebrand 2014 for an extensive review).

If we exclude the specific cases where collaboration is justified by friendship or is considered as a way to escape academic isolation (Medoff, 2003, Acedo et al., 2006, Hamermesh, 2013), team formation is mainly explained by advocating the role of complementarities in researchers' abilities. In a pioneer paper, McDowell and Melvin (1983) linked the raise of co-authorship to the explosion of knowledge in economics. In an academic world where researchers are involved in increasing specialization, co-authorship allows complementarities and appears as an efficient way to improve scientific production. Under alternative presentations, this seminal argument has been developed in a series of contributions. For instance, Piette and Ross (1992) states that authors who work in areas outside of their specialty tend to engage more in co-authorship than authors with close scientific tools. More recently, focusing on Nobel Laureates' pattern of co-authorship, Chan et al. (2015), shows that scientific collaboration may be induced by conceptual complementarities – complementarities that erode through time after repeated interaction.

Another line of research considers the gender dimension of team formation. Boschini (2007) puts forward evidences of gender sorting in team formation. She underlines that the propensity to co-author with a woman is higher for women than for men. Moreover, this propensity gap increases with the presence of women in the field of research. This result is consistent with experimental evidence that the gender composition of teams affects team productivity (Ivanova-Stenzel and Kubler 2011). The fact that women seem to perform worse in case of gender mixed teams added to the under representation of women in research activities may contribute to explain a lower rate of co-authorship in the researchers' female population.

More anecdotal factors may also influence team formation. For instance Ong et al (2015) shows that co-authorship may be affected by the order in which authors are listed on title pages. The authors show that, as authors with earlier last names initials have better visibility, they are therefore more keen to start collaborations.

In these settings, our work aims at following an intuition by Fafchamps et al. (2010) who underlines the matching problem of finding a co-author (see also Hamermesh 2015). When research output depends on ability, they state that collaboration is most likely between authors of a similar level of ability (what they call assortative matching). Collaboration between authors with different abilities can only arise if the contribution of the low skilled author relaxes the time-constraint of his/her co-authors. In this agreement, high skilled authors produce more research while low skilled researchers increase the mean quality of their output.

In order to examine the consequences of this intuition, our paper presents first a simple theoretical model in which authors with different level of ability are randomly matched. Each researcher has to decide if he/she accepts to collaborate or if he/she prefers to work alone. By assumption, collaboration allows saving time on a paper but, according to the characteristics of both researchers or on the assessment rules, the co-authored scientific production may present a lower value than a singled authored paper. Built in a dynamic setting, the model leads to the characterization of an optimal decision rule leading to the choice of collaboration. In a nutshell, three main conclusions arise from the model: first, the higher the ability of a researcher, the higher will be the skills of his/her co-authors. Second, the number of papers written during a given period of time is increasing with the productivity of the researcher. Finally, a talented author should more frequently meet authors willing to accept collaboration and should have more co-authored papers than authors with low productivity.

The second part of the paper provides an empirical assessment of the model's prediction. The purpose of this section is to check the link between the quality of a researcher and the one of his / her co-authors. For this purpose, we built an original database considering all economists with a position in a French university during the year 2004. In 2012, we collected their publication records in order to take into account the scientific work of these academics with at least 8 years of research experience. We then computed the h and g indexes of the academics and their co-authors and, in order to consider the number and the quality of the coauthors in a single one-dimensional variable, we built two meta indexes (hereafter the hh and gg indexes) by reference to the h and the g indexes[3]. More specifically, a hh index with value *i* indicates that the author has at least *i* co-authors with a h index superior or equal to *i*. A high hh index reveals that the author collaborates with a high number of influential co-authors. The gg index of an author is equal to *i* if *i* is the highest value for which the sum of the g-indexes of his/her *i* best co-authors is superior or equal to i^2. The gg index focuses mainly on the influence of the co-authors. A gg index may take a high value when the author collaborates with a high number of influential co-authors or if some of the co-authors have a very high g index. Our prediction is that the hh index (resp gg index) of a given author is explained by the author's h (resp g) index. In other words, the more talented a academic is, the higher will be the number and the quality of his/her co-authors. The database incorporates several control variables in order to eliminate the influence of side effects such as age, gender, reputation, localization of the university. Basic results are consistent with existence of an assortative matching which leads academics of equivalent abilities to work together.

From a methodological point of view, assessing the relationship between the characteristics of an academic with those of his/her co-authors raises two major issues. The first is linked to the excess number of zeros in the data. An important share of academics exhibit academic CVs without any scientific publication and consequently without any scientific collaboration. These zeros request a specific econometric treatment. A second difficulty appears as there is an obvious endogeneity issue in our model: the academic production of a researcher and the one of his/herco-authors are clearly interlinked and the h index (resp g index) of a given author could also be explained by his/her co-author's productivity (as measured by the hh or the gg indexes). These two problems and their technical solutions will be considered in a separated section.

Finally, one paradoxical consequence of our result is to emphasize the positive signaling role of co-authorship. In their contribution, Fafchamps et al. (2010) stressed that co-authorship suffers from a free-rider curse. The imperfect information about the cooperative willingness of a potential co-author reduces the incentive to collaborate. They conclude that interpersonal links that allows getting private information on the cooperative behavior of an academic play an important role in team formation. However, as long as co-authoring appears as a voluntary choice, no long run collaboration seems feasible with an academic free rider. As remarked by Ishida (2009), repeated co-writings between the members of a same team thus constitute a signal of researchers' collaborative behavior. Contrary to what had been presented in our first lines, co-authorship appears here as a positive signal of the researcher's willingness to carry a part of the workload. In this paper, the positive nature of the co-authorship signal is developed one step further. As academics with a lot of influent publications work and create links with a large number of influential co-authors, co-authoring with influent researchers appears as a signaling device of academics' quality. In order to improve the information about researchers' ability, co-authorship should be encouraged.

[3]Meta indexes are often used in bibliometrics. For instance, Schubert (2012) computes a meta index to measure the characteristics of a researcher's network and Tol (2008) proposes a generalized g index to rank groups of researchers.

The paper is organized as follows. The next section presents the theoretical model. Section 2 provides the methodology of the empirical model. Section 3 describes the data and section 4 presents the empirical results. A last section concludes the paper.

Section 1: A simple theoretical model

We consider a population of researchers of dimension one, each researcher being defined by his/her level of productivity q, with q∈[0,1]. At each point of time, Nature randomly matches couples of researchers and gives them the opportunity of a scientific collaboration. Researchers have then to decide if they accept or not this collaboration. When they accept to work together, authors co-write a paper and, after completion of their work, they are available for a new collaboration opportunity. If one of them refuses to collaborate, both will write their next paper alone. Authors make their choice in order to maximize their utility function in continuous time and with infinite horizon.

- The Bellman equation

Denote $V^Q(q)$ the discounted expected utility of a researcher with a productivity level Q (hereafter the Q-researcher) matched with a researcher with productivity q (the q-researcher). For the Q-researcher, the expected utility before the random choice of a potential co-author is $W^Q = \int_0^1 V^Q(q)dF(q)$, where F(q) represents the cumulative distribution of the researchers' productivity.[4]

Let us define as Φ(Q) the subset of the q-researchers who would accept to work with the Q-researcher. The expected utility of the Q-researcher matched with a given q-researcher obeys the following Bellman equation:

$$V^Q(q) = \begin{cases} max\{U(Q) + \beta(Q)W^Q, U(Q,q) + \beta(Q,q)W^Q\}, \forall q \in \Phi(\mathrm{Q}) \\ U(Q) + \beta(Q)W^Q,\ \forall q \notin \Phi(\mathrm{Q}) \end{cases} \qquad (1)$$

Where :

· U(Q) measures the instantaneous utility level of the Q-researcher who writes his/her next paper alone and U(Q,q) his/her utility in case of collaboration with the q-researcher.

· β(Q) ∈]0,1[is the discount rate associated with the time spent by the Q-researcher when he writes a paper alone and β(Q,q) ∈]0,1[the discount rate in case of collaboration.

· The present value of the future collaboration opportunities is $\beta(Q)W^Q$ when the Q-researcher works alone and $\beta(Q,q)W^Q$ in the other case.

According to the second line of Eq. (1), when the q-researcher refuses to co-work with the Q-author, q∉Φ(Q), both authors will write their next paper alone. The Q-researcher gets utility U(Q) and once his/her paper finished, he/she will be faced to a new opportunity of collaboration with an expected

[4]The model assumes an homogenous distribution F() over the set of researchers. This implicitly means that a top tier economist will have the same probability to be matched with another leading economist as a researcher without any scientific activity. This is obviously a heroic assumption made for sake of simplicity. Considering differences in these distribution functions according to the productivity of each researcher would not modify the main insight of the model.

intertemporal value W^Q. In the opposite case, q∈Φ(Q), the q-researcher accepts to work with the Q-author. According to the first line of Eq. (1) the last one has to decide if he/she prefers to work alone (in this case the intertemporal reward is $U(Q)+\beta(Q)W^Q$) or to collaborate, a decision which leads to the expected reward $U(Q,q)+ \beta(Q,q)\ W^Q$.

We assume that the higher the productivity of the Q-researcher, the higher the reward of his/her scientific production (ie$dU/dQ>0$) and the faster the speed of one paper's production (ie :$d\beta/dQ>0$). In case of collaboration, the discount factor $\beta(Q,q)$ is an increasing function of co-autors' productivities, $\beta(Q,0) = \beta(Q)$,$d\beta(Q,q)/dQ>0$ and $d\beta(Q,q)/dq>0$ and utility is increasing with the productivity of the two authors : $U(Q) > U(Q,0)$, $\partial U(Q,q)/\partial Q>0$, $\partial U(Q,q)/\partial q>0$.

- *The cooperation thresholds*

Consider first the case q∈Φ(Q) in which the q-researcher is willing to collaborate with the Q-researcher. Consider the first line of Eq.(1). Given that the first term in the brackets is independent of q, and that the second term is increasing in q, the Q-researcher's optimal strategy consists in selecting a (reservation) productivity level $q_{min}^Q$and in accepting to work with any researcher with productivity above this value. According to this decision rule, the value function $V^Q(q)$ is defined as :

$$V^Q(q) = \begin{cases} V_1^Q(q) = U(Q) + \beta(Q)W^Q, if\ q < q_{min}^Q \\ V_2^Q(q) = U(Q,q) + \beta(Q,q)W^Q,\ if\ q \geq q_{min}^Q \end{cases}, \forall q \in \Phi(\mathrm{Q}) \qquad (2)$$

where q_{min}^Q is implicitely defined by $V_1^Q(q_{min}^Q) = V_2^Q(q_{min}^Q)$, i.e.:

$$U(Q) + \beta(Q)W^Q = U(Q,q_{min}^Q) + \beta(Q,q_{min}^Q)W^Q \qquad (3)$$

In an alternative way, Eq. (3) may be stated as:

$$U(Q) - U(Q,q_{min}^Q) = [\beta(Q,q_{min}^Q) - \beta(Q)]W^Q \qquad (4)$$

Note that W^Q is a sum of positive elements and is therefore positive. If we admit that working with a q-researcher allows to save time when this co-author presents a strictly positive productivity, we must have $[\beta(Q,q_{min}^Q) - \beta(Q)] > 0$ for any $q_{min}^Q > 0$. Thus, Eq. 4 states that a researcher may accept to collaborate on a paper of little interest, U(Q,q_{min}^Q) <U(Q), because collaboration allows to save time in the writing process.[5]

In the general case, it is difficult to define analytically the relationship between Q and q_{min}^Qas defined by Eq. 4. The simulation in the next subsection shows that under rather credible assumptions q_{min}^Q is an increasing function of Q. However, if a Q-researcher may refuse to work with poorly skilled authors, she/he must also realize that highly skilled researchers could refuse to work with her/him. By replication of the same argument, define $\hat{Q}$ as the productivity of the researcher for which the reservation level of productivity is $q_{min}^{\hat{Q}} = Q$. If $q_{min}^{\hat{Q}}$ is an increasing function of $\hat{Q}$, any q-researcher with a productivity level above $\hat{Q}$ will refuse to work with the Q-researcher. Productivity $\hat{Q}$ thus appears as an upper threshold, $q_{max}^Q = \hat{Q}$, for the Q-researcher who will only be able to cooperate with co-authors presenting a productivity in the range Φ(Q) = [q_{min}^Q,q_{max}^Q].

[5] This point is consistent with Medoff (2003) who emphasizes that collaboration doesn't significantly improves scientific output. See also Fafchamps et al. (2010).

Finally, note that collaboration can only appear if some researchers accept to cooperate with less skilled authors. If Q<q_{min}^{Q}, ∀Q, researchers are only willing to co-work with authors more productive than themselves. In this case, any collaboration opportunity would be rejected by at least one of the two researchers. The documented generalization of co-authorship thus implies that Q>q_{min}^{Q}, and thus Q< q_{max}^{Q}, for a majority of researchers.

- *The collaboration range*

With a set of potential collaborators, Φ(Q) = [q_{min}^{Q},q_{max}^{Q}], the value function $V^{Q}(q)$ - Cf. Eq. 1 - may be restated as (note that q_{max}^{Q}=1 when every researcher is ready to accept a scientific collaboration with the Q-researcher):

$$V^{Q}(q)=\begin{cases}V_1^{Q}(q)=U(Q)+\beta(Q)W^{Q}, if\ q<q_{min}^{Q} \text{ or }, q\geq q_{max}^{Q}\\ V_2^{Q}(q)=U(Q,q)+\beta(Q,q)W^{Q},\ if\ q_{min}^{Q}\leq q<q_{max}^{Q}\end{cases}, \forall q\in\Phi(\mathrm{Q}) \quad (5)$$

With:

$$W^{Q}=\int_0^{q_{min}^{Q}}V_1^{Q}(q)dF(q)+\int_{q_{min}^{Q}}^{q_{max}^{Q}}V_2^{Q}(q)dF(q)+\int_{q_{max}^{Q}}^{1}V_1^{Q}(q)dF(q) \quad (6)$$

and where q_{min}^{Q} and $q_{max}^{Q}$are implicitely defined by :

$$\begin{cases}U(Q)+\beta(Q)W^{Q}=U(Q,q_{min}^{Q})+\beta(Q,q_{min}^{Q})W^{Q}\\ U(q_{max}^{Q})+\beta(q_{max}^{Q})W^{q_{max}^{Q}}=U(q_{max}^{Q},Q)+\beta(q_{max}^{Q},Q)W^{q_{max}^{Q}},\end{cases}$$

The threshold $q_{\min}^{Q}$ is defined in the range [0,q_{max}^{Q}] under the necessary and sufficient condition (see Appendice 1):

$$\frac{U(Q)}{1-\beta(Q)}<\frac{U(Q,q_{max}^{Q})}{1-\beta(Q,q_{max}^{Q})} \quad (7)$$

Note that by construction $\mathrm{q}_{\min}^{\mathrm{Q}}$ and $\mathrm{q}_{\max}^{\mathrm{Q}}$ are symmetrical with respect to the first bisector. Thus, when condition (7) is fulfilled, $q_{min}^{Q}<Q<q_{max}^{Q}$, and the Q-researcher will accept to write a paper with any q-researcher with productivity within the range$[q_{min}^{Q}, q_{max}^{Q}]$. In all other cases, the Q-researcher always prefers to work alone.

- *Simulation*

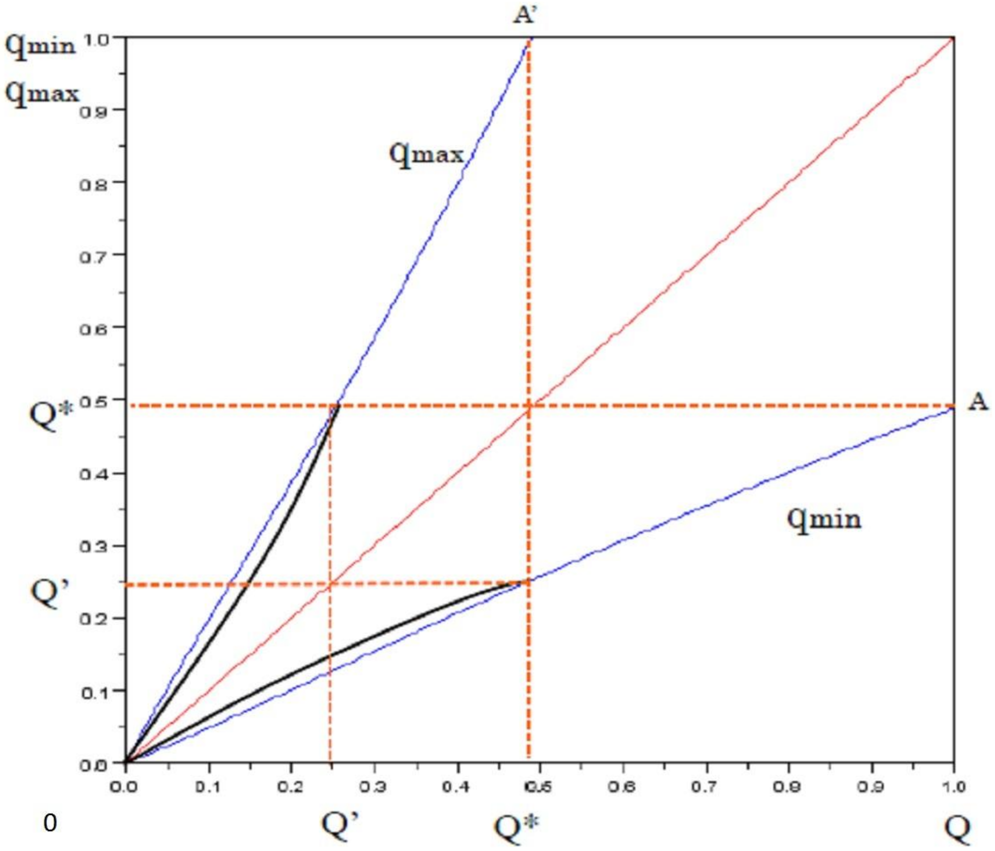
In order to get some intuition about the collaboration range, let us consider the additional assumptions:

- Define as t(.) the time spent by a Q-researcher to write a paper with:

$$\begin{cases}t(Q)=\dfrac{N}{N+Q},\ \text{when the researcher works alone}\\ t(Q,q)=\dfrac{N}{N+Q+q}\ ,\ \text{in case of collaboration}\end{cases}$$

Here, N is a constant that allows weighting the productivity's influence on writing time. For sake of simplicity, we will assume N=1 in the simulation. With this definition of t(), the discount factor respects the usual definition $\beta(-)= e^{-rt(-)}$, where r measures the researcher's discount rate.

- Assume q to be distributed according to a uniform distribution on [0,1], with cumulative distribution $F(\mathrm{q}) = \mathrm{q}$.
- Assume U(Q)=Q, and $U(Q,q)=[aQ^{s/(s-1)}+aq^{s/(s-1)}]^{(s-1)/s}$, i.e.: the joint production function presents a constant elasticity of substitution.[6]



-

Fig. 1

Fig. 1 presents the values (q^Q_{min}, q^Q_{max}) for each Q-researcher under the previous assumptions with a=0.5, s=2, r=0.04. The curves q^Q_{min} and q^Q_{max} indicate the threshold values for each Q-researcher. They are symmetric with respect to the 45° line and computed starting from Q=1 (and therefore q^Q_{max}=1) and deducing the two threshold values for each lower level of productivity.

According to Eq. (5), authors with Q=1 would only accept to work with co-authors with productivity above q^1_{min} = A = Q*. In another way, this implies that any researcher would accept to work with a colleague with a productivity level above Q*. For any Q-author with Q>Q*, the upper threshold takes the value q^Q_{max}=1.

[6]Krapf (2014) uses the same production function to formalize the result of collaboration; Medoff (2007) uses a generalized form of the CES.

The blue curve 0A indicates the q^{Q}_{min} values for any Q-researcher for which q^{Q}_{max}=1. The symetric of this curve with respect to the 45° line - curve OA' - represents the inverse of the previous function. It gives the q^{Q}_{max} values for researchers with productivity in the range [Q′,Q*].[7] As for any researcher with Q<Q*, the upper threshold is less than one, authors become more demanding with their coauthors. The curve deviates from the blue curve OA and the spread between the two thresholds drops. The symmetric black curves in Fig 1.indicates the values (q^{Q}_{min},q^{Q}_{max}) for Q<Q*.

Three properties of the model deserve attention. First, when the two thresholds are non decreasing with productivity Q, the higher the researcher's productivity, the higher will be the skills of his/her co-authors. Second, as the time spent on a paper is decreasing with the productivities of the researcher and of his/her potential co-authors, the number of papers written during a given period of time must be increasing with Q. Highly skilled authors must produce a higher number of papers (coauthored or not). Finally, as long as the Q-researcher's productivity is defined in the interval [0,Q*], the spread [q^{Q}_{min},q^{Q}_{max}], is increasing with Q, thus a talented author should more frequently meet authors prone to accept a collaboration. He should have more co-authored papers than an author with low productivity. For authors with productivity in the range, [Q*,1], the link between productivity and the number of collaborations depends on the relative influence of a higher frequency of match with other researchers and the decrease in the probability that collaboration will be accepted. The number of coauthored papers may increase or slightly decrease with the productivity level of the Q-researcher.

Section 2: Econometric Methodology

In a nutshell, our theoretical model states that a highly skilled author should have more co-authors of better quality. Thus, the econometric part of this paper aims at estimating a relationship such as:

$$(2.1)\quad Q_{i,co-authors} = f\left(Q_i, X_{ij}\right),\ \text{for i=1, .., N and j=1, ..., M.}$$

where Q_i stands for the quality of researcher i, $Q_{i,co\_authors}$ represents the average quality of his co-authors and $X_{ij} = [X_{1ij} \quad X_{2ij}]$ stands for exogenous variable j of researcher i.

Two main issues arise in this case. Firstly, we will apply count data measures of either the individual research productivity or the quality of the co-authors (see next section for a presentation of the data). So we should apply count data econometrics to the former specification. Secondly, the academic production of a researcher is not independent of his co-authors' productivity level[8]. Thus, there is an endogeneity issue in the data which could be address by the Two Stage Residual Inclusion approach (2SRI, Terza et al, 2008).

[7]For instance, note that the Q*-researcher's lower threshold is Q'=q^{Q*}_{min}. Thus, for the Q′-researcher, the upper threshold is $q^{Q'}_{max}$=Q* on the OA' curve.

[8] For example, in the case of a sample of French physicists, Mairesse and Turner (2005) demonstrate that individual productivity is explained by the quality of other researchers belonging to the same research center.

2.1- Count data with overdispersion and Excess zeros: ZIP and ZINB modeling

Poisson regression models provide a standard framework to analyze count data[9]. However, in practice, count data suffer from two major drawbacks: overdispersion and excess of zeros. Overdispersion could stem from unobserved heterogeneity which causes the conditional variance of the sample to be larger than the conditional mean. The most frequently cited approach to address over dispersion is the negative binomial regression model. Another issue in count data modeling is a situation in which the number of zeros in the data exceeds what would typically be predicted by the Poisson distribution. Lambert (1992) has developed the Zero Inflated Poisson (ZIP) model to handle this case. In order to model both unobserved heterogeneity and excess zeros a Zero Inflated Negative Binomial (ZINB) model could be applied to the data (Greene, 1994).

Zero inflated models suppose that the data generating process is different for the sample values equal to zero and those positive[10]. There should be also a distinction between "*structural* zeros" (which are inevitable) and "*sampling* zeros" (which occur by chance)[11]. For example, we may assume that there are two different types of academics in the sample: those who would never collaborate (for ideological reasons, for example) and the others. Among those wishing to work with other economists, some have not collaborated since they have not found a researcher with whom working. Therefore there are two types of zeros among the observed values, but econometricians cannot distinguish between the two types of individuals. Lambert (1992) introduced the ZIP model in which the zeros values are the result of both a Poisson model and a logit decision process.

In the ZIP model there are two different latent variables: C_i the collaboration decision variable of academic i, and Q_i^* the potential quality level of his/her co-author. The observed quality level of the co-authors ($Q_{i,co-authors}$) is then a function of these two latent variables:

$$(2.2) \qquad Q_{i,co-authors} = \begin{cases} Q_i^* & if \quad C_i = 1 \\ 0 & if \quad C_i = 0 \end{cases}$$

The probability function of the quality level of co-authors is then the following:

$$(2.3) \qquad f\left(Q_{i,co-authors}\right) = \begin{cases} p_i + (1-p_i) \times g(0) & if \qquad Q_{i,co-authors} = 0 \\ (1-p_i) \times g\left(Q_{i,co-authors}\right) & if \qquad Q_{i,co-authors} > 0 \end{cases}$$

where $p_i \in [0,1]$ is the probability that academic i will not collaborate (or the probability of a structural zero), and g(.) is the probability function of the parent count model. Excess zeros occur whenever $p_i > 0$. The collaboration decision C_i will depend on a new latent variable C_i^* and it will be modeled with a logistic model:

$$(2.4) \qquad C_i = \begin{cases} 1 & if \quad C_i^* = X_{1ij}^{'}\delta_1 + \varepsilon_i \geq c_i \\ 0 & if \quad C_i^* = X_{1ij}^{'}\delta_1 + \varepsilon_i < c_i \end{cases}$$

[9] See Ridout et al. (1998) for a review.
[10] See Garay et al. (2011) for a detailed analysis of zero inflated models.
[11] Staub and Winkelmann (2013) makes the distinction between *structural* or *strategic* zeros and *incidental* zeros.

whereX_{1ij} are the exogenous variables involved in the decision process, c_i is a threshold value and ε_i is a residual following a logistic density function. Accordingly, the probability of a structural zero is defined as follows:

$$(2.5)\ p_i = \frac{exp\left(X'_{1ij}\delta_1\right)}{1+exp\left(X'_{1ij}\delta_1\right)}$$

A fully parametric zero-inflated model is then obtained once the probability function of the parent count model is specified. If *g(.)* is a Poisson probability function, then we get the ZIP model (Lambert, 1992):

$$(2.6)\begin{cases} g\left(Q_{i,co-authors}, \lambda_i\right) = \dfrac{exp(-\lambda_i)\times\lambda_i^{Q_{i,co-authors}}}{Q_{i,co-authors}!}, \quad \lambda_i > 0 \\ \lambda_i = exp\left(X'_{2ij}\delta_2\right) \end{cases}$$

Where X_{2ij} are the exogenous variables explaining the expected value of the quality level of academic i's co-authors. The mean of the zero-inflated count data model (i.e. the expected value of the quality of the co-authors) and its variance are then:

$$(2.7a)\ \mathrm{E}\left(\mathrm{Q}_{\mathrm{i,co-authors}}/\mathrm{X}_{1\mathrm{i}}, \mathrm{X}_{2\mathrm{i}}\right) = (1-\mathrm{p_i})\lambda_\mathrm{i} = \frac{\exp\left(\mathrm{X}'_{2\mathrm{ij}}\delta_2\right)}{1+\exp\left(\mathrm{X}'_{1\mathrm{ij}}\delta_1\right)}$$

$$(2.7b)\ V\left(Q_{i,co-authors}/X_{1i}, X_{2i}\right) = (1-p_i)\lambda_i(1+p_i\lambda_i)$$

The ZINB model is obtained if the *g(.)* function is a negative binomial distribution function, the new probability function of the quality level of co-authors is then (Garay*et al*, 2011):

$$(2.8) f\left(Q_{i,co-authors}\right) = \begin{cases} p_i + (1-p_i)\times\left(\frac{\phi}{\lambda_i+\phi}\right)^{\phi} & if \quad Q_{i,co-authors} = 0 \\ (1-p_i)\times\frac{\Gamma(\phi+Q_{i,co-authors})}{\Gamma(Q_{i,co-authors}+1)\Gamma(\phi)}\left(\frac{\lambda_i}{\lambda_i+\phi}\right)^{Q_{i,co-authors}}\left(\frac{\phi}{\lambda_i+\phi}\right)^{\phi} & if \quad Q_{i,co-authors} > 0 \end{cases}$$

Where $\alpha \equiv \phi^{-1}$is the dispersion parameter and *Γ(.)* is the gamma function, then the first two conditional moments are defined as follows:

$$(2.9a)\quad E\left(Q_{i,co-authors}/X_{1i}, X_{2i}\right) = (1-p_i)\lambda_i$$

$$(2.9b)\quad V\left(Q_{i,co-authors}/X_{1i}, X_{2i}\right) = (1-p_i)\lambda_i(1+\alpha\lambda_i+p_i\lambda_i)$$

Unobserved heterogeneity is linked to the parameter α. If the coefficient α is different from zero there is unobserved heterogeneity in the data and the binomial model should be used instead of the Poisson model. Unobserved heterogeneity can be tested by a likelihood ratio test on parameter α with the Vuong test. It is worth noting that the X_{2i} variables can be identical to the X_{1i} ones, overlap with X_{1i} or be completely distinct from X_{1i}. The parameters δ_2 and δ_1 can be interpreted respectively as the semi-elasticities of the parent model and the changes in the log-odds of strategic zeros.

The former specification implies that each subject is observed for the same time interval, referred to as the exposure. If different subjects have different exposures (t_i), then the natural logarithm of the exposure must be included as an offset, a covariate with regression coefficient set to 1 in the specification (Rabe-Hesketh and Skrondal, 2005):

$$(2.10)\quad \lambda_i' = \lambda_i \times t_i = \exp\left(X_{2ij}'\delta_2 + \ln(t_i)\right)$$

The parameters of a zero-inflated model are estimated by a full maximum likelihood (ML) framework[12].

2.2 Addressing the endogeneity issue in count data: the 2SRI approach

Instrumental Variables (IV) methods are the most common framework for addressing the endogeneity. In linear models, the IV methodology corresponds to the Two Stage Least Squares (2SLS) which is a two-step procedure. In non linear models, the Two Stages Prediction Substitution (2SPS) approach can be considered as the non-linear counterpart of the 2SLS estimation. 2SPS substitutes the endogenous regressors in the estimated equation with their consistent predicted values obtained in a first stage auxiliary regression. However, Wooldrige (2014) highlights that, when the conditional expectation model is non-linear, then 2SPS approach usually produces inconsistent estimates. He advocates applying the Two Stage Residual Inclusion (2SRI) approach which allows getting consistent estimates of the parameters in the structural regression. Terza et al (2008) provides the formal proof of consistency for the 2SRI approach.

The 2SRI estimator has the same first-stage as 2SPS. However, in the second stage regression, the endogenous regressors are not replaced. Instead, the first-stage residuals of the auxiliary regressions are included as additional regressors in the second-stage estimation. Recently, Geraci *et al.* (2014) extended the 2SRI framework to count data models. They consider the following general non linear model for the conditional mean of the outcome ($Q_{i,co-authors}$):

$$(2.11)\quad E\left(Q_{i,co-authors}/x_i, x_{ei}, w_i\right) = M(x_i\beta + x_{ei}\beta_e + w_i\lambda) = M(x_i\beta + \textstyle\sum_{s=1}^{S}\gamma_s\, x_{sei} + \sum_{s=1}^{S}\xi_s\, w_{si})$$

Where *M(.)* is a known non-linear function and the regressors can now be split up between two different components: $X_{2i} = [x_i \quad x_{ei}]$ where x_i is a set of K exogenous regressors and x_{ei} is a set of S endogenous regressors (either discrete or continuous) possibly correlated with a set of S unobservable confounders latent variables (omitted variables) w_i. Endogeneity of regressors x_{ei} may be modelled by the correlation between the unobserved confounder factors with x_{ei} and $Q_{i,co-authors}$ (Terza *et al.* 2008):

$$(2.12)\quad x_{eis} = r_s(v_i\xi_s) + w_{si} \qquad s=1, ..., S$$

where $v_i = [x_i \quad z_i]$, Z_i is a set of at least S instrumental variables satisfying all the necessary conditions, and $r_s(.)$ is a set of S non-linear auxiliary equations.

The 2SRI estimator is then obtained by estimating the following regression:

$$(2.13)\quad E\left(Q_{i,co-authors}/x_i, x_{ei}, \hat{u}_i\right) = M(x_i\beta + x_{ei}\beta_e + \hat{u}_i\psi)$$

[12]However, misspecified overdispersion in the model would invalidate ML inference but not the quasi-ML inference. Recently, Staub and Winkelmann (2013) proposed a Poisson quasi-likelihood (PQL) estimator that is robust to misspecification of the overdispersion.

Where $\hat{u}_i$ is a set of S estimated residuals of the first stage equation.. Consistent standard errors of the second-stage parameters can be obtained by bootstrap (Wooldrige, 2012).

In count data models, there is no consensus on how to define the residuals. Geraci et al (2014) advocates to compute two different measures: the raw residual ($\hat{u}_{is} = x_{eis} - E[x_{eis}/w_i]$) and the standardized residual ($\hat{u}_{is}^{std} = \frac{x_{eis} - E[x_{eis}/w_i]}{(V[x_{eis}/w_i])^{1/2}}$). If x_{ei} are count data variables, then the first-stage auxiliary regression can be modelled either by a ZIP or by a ZINB model and the two conditional moments can be computed as stated in equations (2.7) to (2.9). The exogeneity of x_{ei} can be tested via a conventional Wald-type statistics for H_0: $\psi_1 = \psi_2 = \cdots = \psi_S = 0$.

Geraci et al. (2014) extends the literature by analyzing the small sample properties of the 2SRI estimators in count data models. They use Monte Carlo simulations in order to study the power of the exogeneity test and measure the bias of the structural coefficients. Their results show that the 2SRI method has good finite sample properties. Their empirical evidence shows that the power of the test is always higher using standardized residuals. Furthermore, applying standardized residuals leads to smaller bias in the endogenous regressors.

Section 3: Data and Sample Characteristics

3.1- The Database

In order to study the relationship between the scientific performance of a researcher and the one of his/her co-authors, we built an original database mixing various sources of data.

We used the "Tableau de classement du personnel enseignant titulaire et stagiaire", economics section, Ministry for Research (2004) to identify all the academics employed by French Universities on December 31th 2004. For each researcher, this official file allowed us to get information about gender, age, academic status (Full Professor or Assistant Professor) and the university assignment during the year of 2004.

As it is now common in the literature, we used Google Scholar citation indexes in order to measure individual research productivity (Bosquet and Combes, 2012 and 2013). At the beginning of 2012, we used the software "Publish or Perish" (PoP, Harzing 2010) to collect each academic CV from Google Scholar[13]. With this CV came information about the set of published papers, and, for each paper, the numbers of citations, the language used and the authors' names. In a second round, we used PoP to compute the h and the g-indexes of each co-author.

The raw data extracted from Google scholar present some important shortcomings. Authors with names identical to first names raise difficult problems of disambiguation. A query "Philippe Martin" would thus be credited indifferently with the work of Philippe M or Martin P. Authors with frequent French last name such as Petit are credited with papers from homonymous researchers. Married women, who use different author names during their academic life, often present

[13] In 2012, the software PoP offered the option to select papers according to specific subject areas. Our data record all papers identified within the option "Business, Administration, Finance, Economics" and "Social Sciences, Arts, Humanities".

underestimated academic resume. In order to avoid these difficulties, we removed from the database the name of any author for which the disambiguation was hazardous. From an initial number of 1830 names in the "*Tableau de classement*", we kept only 1566 researchers[14].

In order to complete the database, we got information about the research topics of each researcher by collecting the JEL codes of the papers included in our database and listed in Econlit. Finally, we used the dataset “Fichier Central des Thèses”[15] to identify the name of the PhD supervisor and the year of the PhD defense. For foreign PhD or unrecorded thesis in this dataset, information was obtained through individual searches on the net.

With this information, we then computed four indexes of productivity for each researcher.

- *Individual Productivity Indexes*

We computed first the authors' h and g indexes in order to have a synthetic measure of the quantity (number of papers) and the quality (number of citations by paper) of the researcher’s academic production. By definition, the h index of an researcher is equal to x if x of his N papers have received at least x citations each, and the others (N-x) papers have received no more than x citations each (Hirsh, 2005). One drawback of this measure is that two different academics may exhibit similar h indexes even if their respective best papers get a very different number of citations. In order to address this limit, Egghe (2006) has proposed the g index as the (unique) largest number such that the top g articles received (together) at least g^2 citations. By definition the g index is at least equal or larger than the h index.

There are large empirical evidences showing that citations have a more important effect on academic earnings than the number of publication. For example, Hamermesh et al. (1982) on a sample of 148 full professors of seven large US universities proves that an increase in the total number of citations by one unit has a larger effect on academic wages than one additional published paper. However, given the load of criticisms addressed to the h and the g indexes (see for instance Bornmann and Daniel 2007), we also built two additional productivity indexes grounded on the quality of the medium in which papers were published (journals, books, or working papers series). Both indexes are built following the methodology used since 2005 by the juries of the French “Concours National d’Agrégation pour le recrutement des professeurs d’économie” leading to the hiring of new full professors in economics (for a description of this nationwide competition, see Combes et al, 2013c). These indexes are defined as the sum of individual score values given to each recorded paper, the scores being computed as the ratio between the weight of the medium of publication and the square root of the number of the paper’s authors.

The first index, denoted LLG index, is computed according to the discrete weight function implemented by the jury of the “Concours 2008” (Levy-Garboua 2008) and grounded on the CNRS (2012) ranking of economic journals. This ranking considered 6 categories of economic journals: the main journals are graded from 1 (the top tier journals) to 4 (the less influential). Two additional categories: MAD (multidisciplinary) and NR (new journals) were also introduced in order to consider multidisciplinary or promising new journals. Following the jury of the “Concours 2008”, we therefore gave 6 points for each publication in a journal graded 1 by the CNRS, 4 for any journal with a grade of 2, 2 when the journal is credited with a grade of 3 and 1 for a grade of 4. Articles published in journals

[14] This choice implies that some of the most productive researchers have been removed from the database.
[15] The “Fichier Central des Thèses” is a French database created in 1968 to identify every thesis being prepared in French universities.

listed in categories MAD (multidisciplinary) and NR (new journals) by the CNRS were credited with a weight of 1. Finally, any publication listed in Econlit but not in the CNRS ranking received a weight of 0.5. This allowed distinguishing non producing researchers from active researchers publishing only in books or in journals with weak economic impact.

The second index, denoted CL, follows the same methodology but considers only papers listed in Econlit journals. In this index, weights are taken from the Combes and Linnemer (2010) ranking of the Econlit journals. In order to make their classification, Combes and Linnemer defined two scores values (Clm or Clh) from 0 to 100 for each of the 1205 journal listed in Econlit. Both scores reflect the same ranking of the journals but the CLh is more selective giving higher weights to the top tier journals and lower one to less influential journals. In our paper, the CL index is built according to the CLm weight[16]. Note that the CL index neglects papers published in books, working papers or journals ignored by Econlit and therefore gives an elitist measure of productivity.

According to this methodology, a researcher with one paper written with one coauthor and published in the American Economic Review (with a CLm weight of 98.1 in the Combes and Linnemer's ranking and listed in category 1 by the CNRS) and a paper published alone in a book would have a LLG index equal to 4,74 (i.e.: 6/1.414 plus 0.5) and a CL index of 69,37 (i.e.: 98.1/1.414). These two additional variables will allow us to perform some robustness checks.

- *Co-authors index*

In order to summarize both the productivity and the number of a researcher's co-authors, we also computed two Meta indexes denoted hh and gg by reference to the h and the g indexes. By definition, the hh index of a researcher will be equal to n if n of his/her N co-authors have at least a h index equal to n, and the other (N-h) co-authors have a h index less than n. In a same way, the gg index will be equal to n if the sum of the g indexes of his/her n best coauthors is superior or equal to n^2 (the square of the rank) and the sum of the g indexes of the n+1 best coauthors is inferior to $(n+1)^2$.

These two indexes aim at giving in a one-dimensional variable a measure of both the number and the quality of a researcher's coauthors. A high gg index indicates that an author works with authors presenting high g indexes (some of the co-authors have published very influential papers). For instance if the g index of these co-authors are g1 = 15; g2 = 12; g3 = 4; g4 = 3; g5 = 2; g6=0, the gg index will be equal to 6 (the sum of the g index for the 6 best co-authors is equal to 36 which is equal to the square of the rank). A high hh index indicates that a researcher presents a high number of productive co-authors (with high h indexes). An author who presents 5 coauthors with the following h index, h1 = 15; h2 = 12; h3 = 4; h4 = 3; h5 = 2 will have an hh index equal to three: only three co-authors present a h higher than 3.

These two Meta indexes present the main advantage of taking into account both the quality and quantity dimension in co-authorship issue. We will also analyze the total number of different co-authors ("NB_COAUTHORS" variable) of an academic i in order to study a potential trade-off between quality and quantity in the choice of coauthors.

- *Control variables*

[16] The index built on CLh weight scheme was too selective to provide conclusive empirical results, especially when it was compared to the h and g conclusions.

For each author the following additional variables have been computed:

- "FEMALE" " is a dummy variable equal to 1 if the individual academic is a woman: this variable allows taking into account a gender effect on the publishing strategy of an individual, if any.
- "AGE" stands for the age of the individual so this variable could control for a kind of generation effect. The influence of this variable on the quality and quantity of co-authors may be ambiguous. Indeed, due to the increasing pressure for publication, "young" researcher may want to write more papers and therefore may be looking for an increasing number of co-workers. On the other hand, young researchers may want to signal their quality by avoiding publishing their first papers with co-authors.
- "NUMBER_YEARS" stands for the number of years since the PhD defense. This variable is a proxy for the academic professional experience. It is worth noting that for each academic, the h, g, hh or gg indexes are computed from the beginning of his/her academic career with a different time exposure for each individual of the sample. In the empirical model we will apply the professional experience variable as an offset one.
- *Academic position*: We also control for the rank position of the individual in the academic career: assistant professor ("MCF") or full professor ("PR") with dummy variables[17]. There are three types of ranking for full Professor: - "Classe Exceptionnelle" (PR_CE), "Première Classe" (PR_1C) and "Seconde Classe" (PR_2C) - and two for Assistant Professor- "Hors Classe" (MCF_HC) and "Classe Nornale" (MCF_CN)-. These variables could also reveal the quality of an academic as the promotion from one position to another one (say from MCF to PR_2C or PR_2C to PR_1C) depends on a national competition where the number and the quality of publications plays a dramatic role.
- *Language of publication:* when papers are published in journals, we identified its language of publication and compute for each researcher the share of their papers published in English (SHARE_GB), in French (SHARE_FR) or in other languages (SHARE_OTHER).
- "WORK_ALONE_ONLY": this dummy variable is equal to 1 if academic i has published at least one paper during his/her academic career and has always refused to co-author a paper.
- "NB_PAPERS" stands for the number of papers listed in Google Scholar for the individual researcher. This variable can be considered as a quantitative measure of the level of production of an individual academic.
- *Papers' quality*: the CNRS classification of economics journals plays a dramatic role in the assessment of economic research in France, both at individual and institutional levels. According to this classification, journals are split in 4 categories (from one to four) with an index decreasing with the quality level of the journals. The top tier journals are thus listed in the CNRS1 category, and the less influent in category 4. Until recently, the CNRS also defined a list of multidisciplinary journals publishing economic papers. We used the classification 2008 to measure papers' quality by splitting up the researchers papers in 7 different categories: CNRS_1 to CNRS_5, ECONLIT_NO_CNRS and MISCELLANEOUS_PAPERS. Variables CNRS_1 to CNRS_4 indicate the number of papers published by a given researcher in the four main categories of journals. CNRS_5 stands for papers published in multidisciplinary (MAD) and new journals (NR). The variable ECONLIT_NO_CNRS records the publications in journals that are referenced by the Econlit database but not by the CNRS and, at least, variable MISCELLANEOUS_PAPERS counts all other items (papers published in journals not referenced by either the CNRS or the Econlit database, working papers and books).

[17] In French academia remains the old « *Maître Assistant* » status which only survives in a few cases. We merge academics with this status in the group of MCF_CN.

- HERFINDAHL_JEL_CODE: The propensity to publish and to engage in co-authorship relationships varies greatly according to the various economic topics. Economic fields of research might exhibit very different rates of co-authorship because they do not belong to the same economic tradition (Witte and Schulze, 2009). Following a now standard methodology (see for instance Fafchamps et al., 2010, Kelchtermans and Veugelers, 2011 or Bosquet and Combes 2013a) we collected the JEL (Journal of Economic Literature) codes of the papers included in our database and listed in Econlit and we identified the economic topics through the letter of the papers' JEL Classification codes.[18]We then computed a Herfindahl index of the different letters' JEL code used by a researcher in order to measure the researchers' degree of specialization: the higher the value of the Herfindahl index, the more the researcher is specialized. A Herfindahl index equal to 1 means that all the publications of the researcher are classified in only one JEL category[19].
- WORK_ALONE_ONLY is a dummy variable equal to 1 if faculty member i has published at least one paper during his/her academic career and he has no co-authors.
- "COWRITE_DR" is a dummy variable equal to one if the academic has written at least one paper with his/her supervisor. If this collaboration reflects the assessment of the student's quality by the supervisor, it may constitute a signal of quality for a young researcher; its effect is expected to be positive.
- Finally we also control for network effects with the two different variables. Firstly, the "UNIVERSITY_NAME" variable is the university assignment of the individual in 2004. This is a dummy variable equal to 1 if the individual works in the assigned university. In our dataset there are 90 different institutions (universities, "Grande Ecole"). We assume that belonging to academic institutions with large economic departments which are recognized nationally and internationally can facilitate the matching with complementary co-authors[20].
- Secondly, we compute the "PhD DEFENDED AT" variable. This variable divides the set of authors into 11 categories according to the academic institution where the individual researcher has defended his PhD. It comprises nine French academic institutions (University of Toulouse 1, Paris 1, Paris 9, Paris 10, the others universities in Paris, Aix-Marseilles, Strasbourg, a group of 12 different "Grande Ecole" institutions and all the others French universities)[21]and two categories listing foreign institutions (the PhD has been defended in either an European country or in the United States).

- *Choice of Instruments*

It is well-known that IV estimators' efficiency relies on the quality of the instruments. Endogeneity issue arises because of unobserved heterogeneity in the data, possibly stemming from unobserved individuals' characteristics, which implies that the dependent variable is correlated with some regressors in the equation. So in our case a good instrument requires two assumptions: (i) to be

[18]We also controlled on the researcher's main field of research considering this field as the one in which a researcher has the highest JEL figure. In the paper at hand, this variable didn't prove to be significant.
[19]We computed a normalized Herfindahl index which means that this variable ranges from 0 (no specialization) to 1 (full specialization).
[20] As recently showed by Bosquet and Combes (2013b), the network effect is better measured at the level of the economic departments rather than universities. However in our dataset, we were unable to obtain this information.
[21]This classification corresponds broadly to the international departments of French academia defined recently by Bosquet *et al* (2013c).

highly correlated with the individual research productivity level and (ii) to be uncorrelated with the quality of co-authors. Getting such an instrument is a challenging issue because a lot of potential variables may explain both individual and co-authors quality.

In this paper we used as instrument the best quality paper published alone by an academic. This variable could be interpreted as a signal regarding the researcher's intrinsic skill level. Again we turned in part to the CNRS classification in order to measure papers quality and we relied on the seven categories already discussed. For example, "BEST_ALONE_CNRS1" is a dummy variable equal to 1 if the researcher has published alone at least one paper in the first category of the CNRS classification.

3.2- Some Descriptive Statistics

In Table 1 are reported some descriptive statistics of our database. In 2004, 28% of French academic economists were women. The average academic is 47 years old and has around 15 years of professional experience[22]. In order to analyze the impact of different generations of academics, we have split-up the professional experience variable into 8 different cohorts. For example, the variable "Cohort64_68" is a dummy variable equal to 1 if the individual has defended his PhD dissertation between 1964 and 1968. Individuals belonging to this cohort have on average 37 years of post PhD experience against 3 years for individuals belonging to the last cohort ("Cohort99_04"). Around 1.3% of academics are members of the first cohort and they are on average 65 year old. For the last cohort, the figures are respectively 13.6% and 34 years old. Around 30% of French academics in our sample have started their career between 1983 and 1993 and 36% between 1994 and 2004. About 65% of academics are Assistant Professors and 35% are Full Professor.

Considering researchers' productivity indexes, 22.5% of French academics have never produced a paper referenced by Google Scholar during their career (table 1). The average French economist has published around 8 papers in his career in the whole sample. This average number of papers is equal to 11 in the sub-sample of publishing academics[23]. There is a huge heterogeneity between academics as regards their production as the number of papers listed by Google Scholar for each researcher ranges from 0 to 157. On average, about 40% of the publishing academics have produced between 1 and 4 papers during their career[24]. As regards the quality dimension of the research outcome, the estimated mean of respectively the h index and g index are 3.25 and 6.02. So according to the h index, French academics have published 3.25 papers on average with 3.25 citations each[25]. Again there is a huge heterogeneity in "quality" among French academics as the h and g indexes range from 0 to 39 and 0 to 84 respectively.

[22] In 2004, the professional experience variable ranges from 0 to 39 years.

[23] The average numbers of papers per year of experience are respectively 0.46 and 0.59. This implies that on average each researcher has published one paper every two year. It is worth noting that this figure corresponds to the new requirement of the HCERES (Haut Conseil de l'Evaluation de la Recherche et de l'Enseignement Supérieur), the national agency for the research evaluation in France, to consider an academic as a publishing one.

[24] 52% of the publishing academics have produce between 1 and 6 papers during their career.

[25] Our results are in line with those obtained by Bosquet and Combes (2013a). Indeed these authors estimated an average g index of 7.25 on a sample of 2,782 French academic economists between 1969 and 2008. One may explain this discrepancy by different kinds of population under study. In our sample, we only take into account academics that have a position in a French university whereas Bosquet and Combes consider also all full-time researchers from the CNRS and the INRA (Institut National de la Recherche Agronomique). These last two kinds of academics do not have teaching loads and may have higher research productivity.

Some simple correlations between the different individual research productivity indicators are reported in table 2. All correlation estimates are significantly different from zero. As expected, the correlation between the h and g indexes (0.95) is almost perfect. It is worth noting that the correlation is also positive and high (0.66) between the citations scores (the h or g indexes) and the Econlit publication scores (the CL_index). Academics who have published more papers on top ranked journals receive more citations for each publication[26]. It is also worth noting that the LLG_index is more correlated with the quality measures of Econlit publication (0.90) than with the citations scores (0.73).

Regarding the quality of publications, only 1.98% of the publishing academics have never published in a journal referenced either by the CNRS or by Econlit (see table 1). French academic economists who develop a research activity choose to publish in journals listed by the CNRS as their career evolution depends mainly on that particular classification: it represents 85.2% of the total number of papers published in a journal. The papers' quality as measured by the CNRS classification is quite low. Indeed around 79% of the total number of published papers belong to the two lowest quality categories (CNRS3 and CNRS4), whereas 13% belong to the CNRS2 category and only 7% to the top ranked category CNRS1[27]. Note that 62.5% of the published papers are drafted in French against 34.8% in English (see table 1). Again there are huge discrepancies between the different generations of academics: more than 75% (respectively 17%) of the papers are drafted in French (English) for academics who started their career before 1968 (our first generation) against only 52.9% (45.1% respectively) for those who started in 1999 (our last generation, see table 5).

We turn now to comment co-authorship indicators. On average each French academic has 4.5 different co-authors in the whole sample and 6.9 co-authors in the sub-sample of academics which engage in co-authorship (see table 1)[28]. 44% of publishing academics of the sample have between 1 and 3 co-authors. It is worth noting that 34.8% of the individuals in the sample have never had a co-author whereas 17.4% have never published a paper by their own. About 15% of the individuals in the sample have written at least one paper with their PhD supervisor. The mean of the hh Meta index which summarizes both the number and the quality of co-authors is 3.2 and the index ranges from 0 to 29. There is over dispersion in the data as the hh index variance is 11.7. A similar result is obtained with the gg index with an even larger range of variation: the mean and variance are equal respectively to 7.4 and 74.1 (see table 1).

In table 3 and table 4 are reported some statistics by gender and by academic position. The most striking feature is that, on average, individual research productivity indexes are lower for women than for men. The lower research productivity level of women could be explained by the fact that they published fewer papers (5 papers on average for a woman against 10 for a man). This gender effect is even more pronounced if individual research quality is measured by the CL index and the LLG index variables[29]. A gender gap can also be noticed in the co-authorships issue: women had on average a fewer number of coauthors of lower quality (see table 3). However, these results will be slightly challenged in the econometric model.

[26]These results are again in line with those obtained by Bosquet and Combes (2012) on a sample of 2,832 French economist academics for the year 2010.
[27]This result might be explained by the fact that the majority of the French economic journals are classified by the CNRS as belonging mainly to the third and fourth category. In the 2008 classification, only one French economic journal is classified as category 2 and none is classified as category 1.
[28]The overdispersion issue is even more important in this case as the variance is equal to twelve times the mean.
[29] This fact could be in part explained by the fact that women in our sample are younger than men: they are 42 years old on average versus 48 year old for men (see table 4). Younger women in our data set could be explained in turn by the increasing participation of women in academia, especially since the eighties (see table 5).

Regarding academic position, there is an overrepresentation of men compared to women (42% against 17%) amongst Full Professors. If on average an academic has 4.5 coauthors, a Full Professor has on average 8 coauthors against 2.6 for an Assistant Professor (see table 4).

As regards fields of economics, the specialization of French academics is quite low. The average value of the Herfindahl JEL Code index is 0.31, but there is a high heterogeneity between French academics (see table 1). Few academics have all their publications in one main field (only 8.2% of the sample). The ranking (by decreasing importance) of the most cited fields are the following: D (Microeconomics), O (Economic Development, Technical Change and Growth), F (International Economics), L (Industrial Organization), E (Macroeconomics and Monetary Economics, B (History of economic thought, Methodology and Heterodox Approaches) and J (Labor and Demographic Economics).

Cohort effects play a critical role in France. According to Raubert and Ursprung (2008), countries whose national academic system has been subject to important institutional changes may be characterized by significant cohort effect in research productivity as it was the case in Germany. A similar argument may be raised in France as can be observed from data reported in table 5. We compute individual productivity and co-authorship indexes per experience year by entry cohort. All variables have a similar pattern and exhibit a structural break in the mid-eighties. For example, the number of co-authors per year remains quite stable for researchers belonging to the first four cohorts, then there is a huge increase for individuals entering in the academic career in the mid-eighties (cohort84_88). Since then there has been a steadily increase in the number of co-authors per year. A similar evolution is obtained for individual productivity indexes (h, g, CL_index and LLG_index). Younger cohorts are more productive and they engage more in co-authorship activities.

Hereafter, we use the best paper published alone as the instrumental variable in the paper. Some descriptive statistics on this instrument are reported in the bottom part of table 1. As expected the number of researchers publishing alone declines when the quality level of journals increases. Indeed there are around 6% of the 1566 academics that have succeeded to publish alone at least one paper in the CNRS1 category, 9.5% in the CNRS2, 27% in the CNRS3 and 11% in the CNRS4.

Section 4: Empirical Results and Sensitivity Analysis

Studying the relationship between the characteristics of a researcher and those of his/her co-authors is tricky as individual research productivity should be an endogenous regressor: the quality of a researcher's publication depends somehow on the quality of his co-authors[30]. So in order to evaluate the effect of the level of its own research productivity on the co-authors quality, this endogeneity bias should be addressed. We start by commenting the results of the 2SRI estimations and then we will check the robustness of our results.

[30]This relationship has been put forward by various empirical studies. For instance, Azoulay et al. (2010) shows that co-authors of a superstar suffer a lasting 8 to 18% decline in their quality-adjusted publication output following his death. Chung et al (2009) states that co-authored papers are cited more frequently, and that in the case of asymmetric partnerships collaborating with a higher quality author seems to pay off. Recently, Bosquet and Combes (2013a) finds that an author who increases on average his/her number of co-authors from two to three is cited 53.4% more and his/her g index will be 41.8% higher.

4.1 Empirical Results: IV estimation outputs

2SRI is a two-step approach. In the first step, the endogenous regressors are modelled with the exogenous regressors and the instrumental variables in order to compute the standardized residuals. In the second step, the structural model is estimated and the standardized residuals are included as additional explanatory variables in the regression.

First-Stage empirical results: exogenous determinants of research productivity measures

Both research productivity variables (the h and g indexes) are explained by some exogenous demographic variables, a gender effect and finally the instrumental variable. As there is over dispersion in the data (see table 1), ZINB modeling has been applied to research productivity indicators. Results are reported in table 5[31].The Vuong test compares zero-inflated models with an ordinary Poisson regression model (the null hypothesis). As the Vuong test rejects the null hypothesis whatever the productivity measure applied, a zero-inflated model is better than a Poisson regression[32] (see table 6 column 1). Furthermore, the likelihood ratio tests show that the null hypothesis of no unobserved heterogeneity is rejected at the 1% level whatever the productivity variable applied (again in the case of the h index, the statistic is equal to 589.77 with a p-value of 0%). So these two tests indicate that individual research productivity measures should indeed be estimated with a ZINB model.

Regarding the determinants of the decision to undertake some research activity, the inflate coefficients of the two productivity measures provides similar conclusions. The age and the gender variables have a significant positive effect on the log odds of an inflated zero, and the dummy variable "Working_alone_only" has a non-significant effect. In the case of the age variable, the estimate ranges from 0.16 to 0.22. For example, regarding the h index results, an additional one year old will increase the log odds of an inflated zero by 0.22 (see the inflate model in table 6 column 1). In other words, the older the researcher, the more likely is the fact that not having published a paper is a deliberate decision. There is a gender effect as being a woman increases the log odds of not publishing by 2.33 in the case of the h index[33]. We have also introduced in the regression some cohort dummy variables in order to model life cycles in research productivity (see below for further details). In the case of the h index specification, five out of the six cohort dummies are significant against only three out of 6 for the g index case.

We turn now to comment the results for the parent model of the research quality measures based on citation scores. The level of these individual research productivity measures may be interpreted as the effort allocated by each academic to research. The number of year of professional experience (the "number_years" variable) is the offset variable as each researcher has different time exposure. The evidence regarding the existence of a gender effect is mixed as the dummy female variable is significant in the h index specification but not in the g index specification (see table 6 columns 1 and 2). Being a woman cuts the expected h index estimate by 14%. As expected the age variable has negative impact on the research productivity of an individual as the IRR estimates are below 1 in both cases. So younger economists tend to be more productive: for example, the expected

[31]According to Staub and Winkelmann (2013) identification of all parameters in a ZINM model is achieved if at least one variable in X_2 is not included in X_1.

[32]For example, in the case of the h index, the statistic is equal to 3.58 with a p-value equal to 0.2%.

[33]The estimate is lower for the other productivity measure, but it remains significant.

change of h index if an individual has an additional year old is -0.012 (ln(0.988491))[34]. However, the age effect on the research productivity is rather fragile as the age variable is significant only at the 10% level in the case of the h index. So the age effect doesn't seem to be an important determinant of French academics research productivity.

Research productivity may also vary across historical times because of institutional changes (Raubert and Ursprung, 2008 and Hamermesh, 2015). We include cohort dummies in the specification in order to capture this effect. There is strong evidence for the presence of life cycles in research productivity in our sample of French academics. The reference dummy cohort is the last one (i.e. academics that have defended their PhD after the year 1999), all the seven cohort dummies are significant whatever the measure of co-authorship applied. Most of these dummy variables are significant at the 1% level. The IRR estimates of the cohort dummies are increasing over time implying that the coefficients are decreasing over time as expected (see table 6 columns 1 and 2). For example, an academic having defended his PhD between the years 1969-1973 exhibits a 41% lower h index than that of the reference group. If the PhD has been defended between during the period 1989-1993, the decrease in the h index is only around 22%. The reason of this phenomenon should be very similar to the one put forward by Rauber and Ursprung (2008) in the case of Germany. Over the last forty years, French academics have been increasingly exposed to the Anglo-Saxon research tradition that rewards researchers according to their own research productivity efforts. Furthermore, bibliometrics has become increasingly significant in evaluating individual researchers in French academic system since the nineties (Académie des Sciences, 2011).

Most importantly, our selected instrumental variables are significant and they have the expected signs on research productivity. Indeed the coefficients of the "best alone publication" dummies are increasing according to the quality of journals whatever the measure of research productivity applied. For example, in the case of the h index, the estimated IRR coefficients for "Best_alone_CNRS1" and "Best_alone_CNRS4" are respectively equal to 4.87 and 1.86. These results imply that an individual that has published alone in a journal belonging to the CNRS category 1 (the best quality) has on average a 487% increase in his expected productivity whereas the expected increase in productivity will be solely of 86% if his best publication alone is the CNRS category 4 (the lowest quality)[35].

Second-stage empirical results: IV estimation results of the determinants of co-authorship

In Table 7 are reported the empirical results of the determinants of co-authorship by the 2SRI methodology. We analyze the co-authorship decision both in terms of quality with the number of citations (the hh and gg indexes) and in terms of quantity (the number of co-authors). According to Geraci et al (2014)'s Monte Carlo results, for applied research the best model is a ZINB or a ZIP model with standardized residuals with non-corrected standard errors of the parameters. In every specification the standardized residual variables are significant. Furthermore, the four Wald tests always reject at the 1% level the null hypothesis of exogeneity of the individual research productivity variables. So the 2SRI methodology must be implemented in order to address the endogeneity issue in the data (see the bottom part of table 7, columns 1 to 4).

[34]We have also introduced in the specification the age squared variable, but this variable was never significant.
[35]The reference here is not having published a paper.

Once a network effect taken into account, the likelihood ratio test no more rejects the null hypothesis of no unobserved heterogeneity in the case of the hh index. In this case the best model is the ZIP model. If the dependent variable is the gg index, the selected model is the ZINB: indeed the likelihood ratio test the null of no heterogeneity is always rejected and the Vuong test also rejects the null hypothesis of a Poisson model (see table 7 columns 3 and 4).

Regarding the determinants of the collaboration decision, the inflate coefficient of the individual research productivity variable has always a negative and significant effect (see inflate model in table 7 columns 1 to 4). For example, when the co-authorship quality is measured by the hh index, the inflate coefficient of h suggests that for each unit increase in h the log odds of an inflated zero decrease from a minimum value of 3.37 up to a maximum of 3.41 (table 7, inflate model columns 1 and 2)[36]. In other words, the higher the research productivity level of an academic is, and the less likely is the decision of never collaborate. More productive academics tend to engage more in co-authorships. The variable "Age" has a negative but never significant effect on the decision of not co-writing papers whatever the research productivity measure. Estimates of a gender effect on the collaboration decision are never significant. Again, some cohort dummies are also significant in the decision of engaging in co-authorship.

We turn now to comment the results for the parent model of the quality measures of co-authorship based on citation scores. Again, the professional experience variable is used as the offset variable. The h index (respectively the g index) of an individual has a positive and significant effect at the 1% level effect on the hh index (respectively the gg index). The IRR coefficient for the h index ranges from 1.109 to 1.126 (table 7, columns 1 and 2), implying that if the h index increases by 1%, then the expected value of co-authors quality level increase will range from 10.9% to 12.6%. In the case of the gg index, estimates are lower (around 2.1%) but remain significant (columns 3 and 4). These results confirm our theoretical model stating that the co-authors research quality of an economist depends on his own research quality.

As expected the age variable has a significant and negative impact on the quality of the co-authors at the 1% level. The average of the IRR coefficient estimate is around 0.985: thus the expected decrease of quality of co-authors if an individual has an additional year old is 0.015 (- ln(0.985)). Again the existence of a gender effect is fragile. The gender variable is only significant with the hh measure of co-authorship quality. In this case, being a woman reduces the expected quality of co-authors by around 15%. Publishing a paper with his/her PhD supervisor has a positive and significant effect (at the 1% level) on the quality of the future co-authors. The estimated IRR for "cowrite_dr" is around to 1.26, implying that an individual that has published at least one paper with his/her PhD supervisor has on average a 26% increase of the expected quality of his co-authors. This result could be explained by a reputation effect. Co-writing with his/her student may signal an implicit recognition by a supervisor of the quality of this student. For a young researcher in French academia, publishing with his PhD supervisor may be a signal of efficiency. The rank position has also a positive and significant effect on the quality of co-authors. Compared to an assistant professor at the standard level (MCF_CN), being a full professor improves the hh index (respectively the gg index) by 20.4% (resp. 27.2%) for a full professor at the highest level (PR_CE) and by 13.25% (12.9%) for a full professor at

[36] As regards the gg index, the estimated coefficients are around -0.75.

the first level (PR_2C). Oddly enough, being an assistant professor at the last level (MCF_HC) reduces by around 12% the hh index and by 11% the gg index but coefficients are not significant[37].

As in the case of the individual research productivity, there are significant life-time cycles in the quality of co-authors. Indeed the IRR coefficients of the cohort dummies are increasing over time either with the hh index as dependent variable or with the gg index. Thus younger cohorts are collaborating with better quality co-authors. As most cohorts dummies are non-significant in the inflate model, the life-time cycles should have a very small impact on the probability of never collaborating with other economists however, they have a positive effect over time on the expected value of the co-authors quality.

Interesting results are obtained regarding the link between the number of papers and their quality and the co-authorship decision. As expected the CNRS classification of journals has a huge effect on the selection of the co-authors, but surprisingly the effect of journals quality goes in the reverse direction. Indeed, the IRR coefficient of the number of papers published in top quality level journals (CNRS1) is significant but below one whatever the measure of the quality of co-authors applied (hh index or gg index), implying that the number and the quality of co-authors decrease with the increase in the number of papers published in the CNRS1 category. On the contrary, the IRR coefficient is significant (only in the case of the gg index) and above one (in all cases) as regards papers published in the CNRS2 category, implying an increase in the number and in the quality of co-authors in this case. In general, the number of papers published in the others CNRS categories (CNRS3, CNRS4, and CNRS5) have a non-significant effect[38].This result could be interpreted in the following way: academics may have a strategic behavior when dealing with the collaborating issue. To successfully publish in journals classified as CNRS2, an academic decides to collaborate with a higher number of coauthors or with more productive coauthors[39]. On the contrary, academics who have a sufficient level of research productivity to publish in top ranked journals (here CNRS1 journals) prefer working with less coauthors in order to reap a higher prestige of the publications.[40]

Publication language also plays an important role in the empirical model. This result is not surprising because in general English journals have a higher impact factor than French ones in all rankings. The IRR coefficient of the share_gb variable is larger than one and highly significant: an increase by 1% in the share of paper written in English raises the hh index (respectively the gg index) by 32.4% (48.9%). Two arguments may explain this relationship. On a first hand, French economists may be interested in developing international collaborations in order to produce papers in better English and to publish in more influential journals (for the influence of English proficiency on research performance, see Olney 2015). On the second hand, proficiency in English is a necessary condition to meet efficient co-authors and engage collaboration with more authors of better quality. In both cases, publishing in English contributes to increase the number and the quality co-authors.

Finally economic fields are also an important determinant in the co-authorship issue. As expected, an increase in the level of economic specialization, that is an increase in the JEL Herfindahl

[37]This result might stem from the fact that the transition from MCF_CN to MCF_HC should depend more on the age of an individual than on the quality of his publications in the French academic system.

[38]If the dependent variable is the gg index, then the number of papers published in the CNRS3 category has also a significant effect.

[39]Another interpretation is the following: French academics who are trying to signal their skills ability in research by publishing in English journals ranked mainly in CNRS2 classification decide to collaborate with more coauthors of better quality.

[40]Authors who publish in the top tier journals may also have difficulties to find efficient co-authors; this argument is consistent with the conclusions of the theoretical model.

index, tends to reduce the number and the quality of co-authors. Being interested by a wide variety of topics allows an easier finding of coauthors.

The hh and the gg indexes measure simultaneously the number and the productivity of a researcher's coauthors. We turn now to a strict quantitative measure, namely the total number of co-authors. In column 5 of table 7, the number of co-authors is explained by the h index whereas in column 6 it is explained by the g index. The Wald tests again reject the null hypothesis of exogeneity of the h and g indexes. Results are relatively similar to those obtained with the quality measure of co-authorship except for the individual research productivity variable.

As regards the number of co-authors, there is no difference between male and female academics. Again, there is a significant age effect: younger economists tend to collaborate with more co-authors. The number of co-authors is also marked by significant cohort effects as all cohort dummies are significant. The IRR estimates are increasing over time: younger cohorts are collaborating with more co-authors. Finally, the most striking result is that the h or g indexes estimates are no more significant (see table 7 columns 5 and 6). Thus individual research productivity is mainly a determinant of the quality of co-authors and has no effect on the quantity measure.

4.2 Sensitivity Analysis

We run a sensitivity analysis by relying on EconLit publication scores as a measure of the individual research productivity, namely the CL and LLG indexes. So we test for the robustness of our results to different measures of individual research productivity. Our comments emphasize mainly the impact of these indexes on the quality and the number of co-authors. Again we use the *best alone* publication as the instrumental variable. The first stage estimates are reported in table 6 (columns 3 and 4) and the 2SRI results are reported in table 8. In the first stage estimation, we have applied a Heckman selection model as the individual research productivity measures are no more count data. In both cases, the Wald test of the null of independent equations is rejected at least at the 5 % level: this clearly supports the Heckman selection equation in our data. Results are very similar with these two different measures of productivity. Most importantly, our instruments are highly correlated with the productivity measures (see table 6)[41].

According to the Wald tests, the null hypothesis of exogeneity is always rejected (see table 8). Thus individual research productivity measured by the quality of Econlit publications is also endogenous to the number and the quality of co-authors. Once this bias corrected by applying the 2SRI methodology, quality of Econlit publications and citation scores produce similar results as regards the determinants of co-authorship. There is a significant gender effect: being a woman reduces by about 10% (respectively 3 or 4%) the quality (respectively the number) of co-authors (see table 8). The age variable has still a negative and significant effect on both the quality and quantity of co-authors. Life-time cycles are also significant: again as the cohort dummies estimates are increasing over time, younger economists are collaborating with a higher number of better quality co-authors. The specialization index and the language of the publications have the same impact as already estimated. The most important result is that the CL and LLG indexes have a positive and significant

[41]The main differences between individual research productivity measured either by Google scholar citations or by publication scores rely on the gender effect and to a lesser extent on the cohort effect. Being a woman reduces significantly both the CL and the LLG indexes. Furthermore the dummy cohort variables are less significant with the LLG index. Therefore, a cohort effect seems to be at work mainly with citation scores index.

effect on the co-authorship variables[42]. Thus the conclusion that more productive academics will collaborate with higher quality co-authors is a robust one.

5. Conclusion

This paper studies the matching process between researchers writing collective papers. It focuses on the links between researchers' productivities in a team of co-authors and shows that authors tend to work with colleagues endowed with similar abilities.

In a first step, our theoretical model states that if collaboration allows increasing the quantity and the quality of the academic production, efficient researchers should have more co-authors than less productive ones and these coauthors should be themselves of higher quality. In a second step, we present an empirical check of this theoretical prediction by estimating the relationship between the researchers' h index (respectively g-index) and a meta-index synthesizing the number and the efficiency of his/her co-authors. Our main finding is that there is a positive relationship between the h (resp. g) index of a researcher and our hh (resp. gg) Meta index. This result sheds a new light on collaboration in academic activities. While co-writing is generally seen as the way for low skilled researchers to increase the quality and the quantity of their research output, our paper shows on the opposite that the quality of his/her co-authors constitutes a signaling device for the quality of a researcher.

Others factors appear determinant in the decision of co-authoring a paper. There is a significant gender effect: being a woman has no impact on the probability of never collaborating with other economists but it decreases both the quality and the quantity of co-authors. As expected the academic position of an individual has strong impact on the expected quality of co-authors: Full professors tend to collaborate more with higher productive co-authors. Life-time cycles are also an important determinant of the co-authorship issue. We have demonstrated that younger cohorts of French academics are collaborating with more co-authors who are also more productive. This result has a very important economic policy implication. National public agencies should apply different evaluation criteria according to the time cohort of the academic. Finally, it is worth noting that publishing with his/her PhD supervisor contributes to increase the quality of future co-authors and therefore may be seen as a signal of quality in French academia.

In order to be fully conclusive, additional variables should be considered to analyze a wider dimension of the publication activity. For instance, the size of the institution hiring the researcher, the influence of the academic network or of the research topics on the academic fellows' resume should be also considered to evaluate the robustness of our results. This is let for future research.

[42] For example, if the CL index increases by 1% then the quality of co-authors (the hh index) will be increased by 0.16% and the number of co-authors by 0.25%.

References

Académie des Sciences (2011): «Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs », Rapport remis le 17 janvier 2011 à Madame la Ministre de l'Enseignement Supérieur et de la Recherche. Institut de France, Académie des Sciences, Paris, France.

Acedo, F.J., Barroso, C., Casanueva, C., and J. L. Galan, (2006),"Co-Authorship in Management and Organizational Studies: An Empirical and Network Analysis", Journal of Management Studies, 43(5), pp. 957-983.

Azoulay, P., Graff Zivin, J.S., and J. Wang, (2010), "Superstar extinction", Quarterly Journal of Economics, 125, pp. 549–589.

Beaver, D. B. and R. Rosen (1978), "Studies in Scientific Collaboration", Part I. Scientometrics 1, pp. 65-84.

Bidault, F. and T. Hildebrand, (2014), ".The distribution of partnership returns: Evidence from co-authorships in economics journals", Research Policy, 43, pp. 1002–1013

Bornmann, L. and H-D Daniel, (2007), "What do we know about the h index?", Journal of the American Society for Information Science and Technology, 58(9), pp. 1381–1385.

Bosquet, C. and PP. Combes, (2012), «Un panorama de la recherche française en économie comparant les approches Google Scholar et Econlit», Revue d'Economie Politique, 122, pp. 477-545.

Bosquet C. and PP. Combes, (2013a), « Are academics who publish more also more cited ? Individual determinants of publication and citation records», Scientometrics, 97, pp. 831-857.

Bosquet C. and PP. Combes, (2013b). « Do large departments make academics more productive? Agglomeration and peer effects in research», CEPR Discussion Paper n° 9401.

Bosquet C., Combes PP. and C. Garcia-Penalosa (2013c): "Gender and competition: evidence from academic promotion in France", CEPR Discussion Paper n° 9711, October.

Boschini, A. (2007), "Is team formation labor neutral ? Evidence from co-authorship patterns", Journal of Labor Economics, 25(2), pp. 325-365.

Bruno, B. (2014), "Economics of co-authorship", Economic Analysis and Policy, 44 (2), pp.212–220.

Card, D. and S. DellaVigna, (2013), "Nine Facts about Top Journals inEconomics," Journal of Economic Literature, 51(1), pp. 144-161.

Cardoso, A. R., Guimaraes, P. and K. F. Zimmermann, (2010), "Trends in Economic Research: An International Perspective", Kyklos, 63(4), pp. 479-494.

Chan, H. F., Önder, A. O. and B. Torgler, (2015), "The First Cut is the Deepest: Repeated Interactions of Coauthorship and Academic Productivity in Nobel Laureate Teams", CREMA Working paper 2015-04. http://www.crema-research.ch/papers/2015-04.pdf

Chung, K.H., Cox, R.A.K. and K.A., Kim, (2009). On the relation between intellectual collaboration and intellectual output: evidence from the Finance Academe. Quarterly Review of Economics and Finance 49, pp. 893–916.

Combes PP. and L. Linnemer (2010), « Inferring Missing Citations: A Quantitative Multi-Criteria Ranking of all Journals in Economics". WP halshs-00520325.

Combes, PP., Linnemer L. and M. Visser (2008), "Publish or Peer-rich? The Role of Skills and Networks in Hiring Economics Professors", Labour Economics, 15, pp. 423–441 (2008).

Ductor, L., (2014), "Does co-authorship lead to higher academic productivity ?", Oxford Bulletin of Economics and Statistics, 77(3), pp. 385–407.

Durden, G. and Perri, T. (1995). "Coauthorship and publication efficiency", Atlantic Economic Journal, 23 (1), pp. 69-76.

Egghe, Leo (2006), "Theory and practise of the g-index", Scientometrics, 69(1), pp. 131-152.

Fafchamps, M., Goyal S. and M. J. van der Leij, (2010), "Matching and Network Effects," Journal of the European Economic Association, MIT Press, 8(1), pp. 203-231, 03.

Garay, A., M., Hashimoto, E. M. Ortega and V. H. Lachos, (2011). On estimation and influence diagnostics for zero-inflated negative binomial regression models. Computational Statistics and Data Analysis, 55, pp. 1304–1318.

Geraci, A., Fabbriand D. and C. Monfardini (2014). "Testing exogeneity of multinomial regressors in count data models: does two stage residual inclusion work?", Department of Economics, Università di Bologna, Working Paper DSE n° 921.

Greene W. (1994): "Accounting for excess zeros and sample selection in Poisson and Negative Binomial regression models", Department of Economics, Stern School of Business, New York University.

Hamermesh, D., S., (2015), "Age, cohort and co-authorship", NBER Working Paper, n° 20938, February.

Hamermesh, D., S., (2013), "Six decades of Top Economics publishing: Whoand How?", Journal of Economic Literature, 51(1), pp. 162-172.

Hamermesh D. S., Johnson G. E. and Weisbrod B. A., (1982), "Scholarship, Citations and Salaries: Economic Rewards in Economics", Southern Economic Journal, 49, pp. 472-81.

Harzing. A.-W., (2010), "The Publish or Perish Book", TarmaSoftwareResearch Pty Ltd, Melbourne, Australia, 266 p.

Hirsch, J. E. (2005). "An index to quantify an individual's scientificresearch output", PNAS, 102(46), pp. 16569-16572.

Hollis, A., (2001), "Co-authorship and the output of academic economists", Labour Economics, 8, pp. 503-530.

Hudson, J. (1996), "Trends in Multi-authores Papers in Economics", The Journal of Economic Perspectives, 10 (3), pp. 153-158.

Ishida, J. (2009), “Incentives in academics: Collaboration under weak complementarities”, Labour Economics, 16, pp. 215-223.

Ivanova-Stenzel, R. and D. Kubler, (2011), “Gender Differences in Team Work and Team Competition”, 2011, *Journal of Economic Psychology*, 32, pp. 797–808.

Johnson, D., (1997), “Getting noticed in economics: the determinants of academic citations”, The American Economist, 41(1), pp. 43-52.

Kelchtermans S. and R., Veugelers, (2011), “The great divide in scientific productivity: Why the average scientist does not exist”. Industrial and corporate Change, 20(1), pp. 295-336.

Kraft, M. (2014), “Age and complementarity in scientific collaboration”, Empirical Economics, DOI 10.1007/s00181-014-0885-8.

Laband, D.N. (1987). “A Qualitative Test of Journal Discrimination Against Women”, Eastern Economic Journal, 13, pp. 149-154.

Laband D., and R. Tollison, (2000), "Intellectual Collaboration", Journal of Political Economy, 108, pp. 632-662.

Lambert D. (1992), “Zero-Inflated Poisson regression, with an application to defects in manufacturing”, Technometrics, vol 34, pp. 1-14.

Lee S. and B. Bozeman, (2005), The Impact of Research Collaboration on Scientific Productivity, Social Studies of Science, 35, pp. 673-702.

Lévy-Garboua, L. (2008), « Rapport sur le premier concours d'agrégation de l'enseignement supérieur pour le recrutement de professeurs des Universités en Sciences Economiques (Année 2007-2008) », Revue d'économie politique, 5(118), pp. 603-23.

Liebowitz, S. and J. Palmer, (1983), “Assessing Assessments of the Relative Quality of Economics Departments”, Quarterly Review of Economics and Business, 28(2), pp. 88-113.

McDowell, J.M. and M. Melvin, (1983), “The Determinants of Co-Authorship: An Analysis of the Economics Literature”, The Review of Economics and Statistics, 65(3), pp. 155-160.

Mairesse, J., Turner, L., (2005), “Measurement and Explanation of the Intensity of Co-Publication in Scientific Research An Analysis at the Laboratory Level”. NBER working paper series no. 11172, National Bureau of Economic Research, Cambridge.

Medoff, M.H. (2003), “Collaboration and the quality of economic research”. Labour Economics, 10, pp. 597-608.

Medoff, M.H. (2007), “The Input Relationship BetweenCo-Authors in Economics. A Production Function Approach”, The American Journal of Economics and Sociology, 66(2), pp. 289-308.

William W. Olney, (2015), “English Proficiency and Labor Market Performance: Evidence from the Economics Profession”, Department of Economics Working Papers, Department of Economics, Williams College, http://web.williams.edu/Economics/wp/OlneyEnglishProficiency.pdf

Ong, D. Chan, H. F., Torgler, B. and Y. Yang, (2015), “Endogenous Selection into Single and Coauthorships by Surname Initials in Economics and Management”, SSRN: http://ssrn.com/abstract=2557365

Petry, G. H. (1988). "A Statistical Analysis of Worldwide Co-authorship Relationships in Scholarly Journals of Business", Journal of Economics and Business 40(2), pp. 169-176.

Piette M. J., and K. L. Ross (1992), “An Analysis of the Determinants of Co-Authorship in Economics”, Journal of Economic Education, 23, pp. 277-283.

Rabe-Hesketh S. and A. Skrondal, (2005), Multilevel and Longitudinal Modelingusing Stata, published by Stata Press Publication, StataCorp LP, CollegeStation, Texas USA.

Rauber M. and H. W. Ursprung, (2008), “Life cycle and cohort productivity in economic research: the case of Germany”, German Economic Review, vol 9, issue 4, pp 431-456.

Ridout M., Demétrio C. and J. Hinde (1998), “Models for count data withmany zeros”, Proceeding of the XIXth International Biometric Conference, CapeTown, December, Invited Paper, pp. 179-192.

Schubert, A., (2012), "A Hirsh-type index of co-author partnership ability",Scientometrics, 91(1), pp. 303-308.

Staub E. K. and R. Winkelmann, (2013), “Consistent estimation of zero-inflated count models”, Health Economics, 22(6), pp. 673-6886.

Stefaniak, B., (1982), “Individual and Multiple Authorship of Papers in Chemistry and Physics”, Scientometrics, 4, pp. 331-337.

Stephan, P. E., (1996), "The Economics of Science," Journal of Economic Literature, 34(3), pp. 1199-1235.

Terza J. V., Basu A. and P. J. Rathouz, (2008), “Two-Stage residual inclusion estimation: Adressing endogeneity in health econometric modeling”, Journal of Health Economics, 27, pp 531-543.

Tol, R.S.J., (2008), “A rational, successive g-index applied to economics departments in Ireland”, Journal of Infometrics, vol. 2, pp. 149-155.

Witte S., and G. G. Schulze, (2009), “Do Birds of a Feather Flock together? Hiring Patterns of academic economists in Germany”, Mimeo University of Saarlandes, Germany.

Ursprung H. W., and M. Zimmer, (2007), “Who is the Platz-Hirsh of the German Economics Profession? A citation analysis”, Journal of Economics and Statistics, 227(2), pp. 187-208.

Wooldridge J. M., (2014), “Quasi-Maximum Likelihood Estimation and testing Nonlinear Models with Endogenous Explanatory Variables”, Journal of Econometrics, vol 182(1) pp. 226-234.

Wuchty, S., Jones, B. F. and B. Uzzy (2007), “The increasing Dominance of Teams in Production of knowledge”, Science, 316(5827), pp. 1036-1039.

Zitt, M., E. Bassecoulard and Y. Okubo (2000), "Shadows of the Past in International Cooperation: Collaboration Profiles of the Top Five Producers of Science", Scientometrics, 47, pp. 627-657.

Appendice 1.

Let us define functions Π(Q,x) and Θ(Q,x) as:

$$\Pi(Q,x) = U(Q) + \beta(Q)\left[\int_0^x V_1^Q(q)dF(q) + \int_x^{q_{max}^Q} V_2^Q(q)dF(q) + \int_{q_{max}^Q}^1 V_1^Q(q)dF(q)\right]$$

$$\Theta(Q,x) = U(Q,x) + \beta(Q,x)\left[\int_0^x V_1^Q(q)dF(q) + \int_x^{q_{max}^Q} V_2^Q(q)dF(q) + \int_{q_{max}^Q}^1 V_1^Q(q)dF(q)\right]$$

Given the threshold q_{max}^Q, the value q_{min}^Q appears as the implicit solution of:

$$\Pi\left(Q,q_{min}^Q\right) = \Theta\left(Q,q_{min}^Q\right)$$

with:

$$\begin{cases} V_1^Q(x) = U(Q) + \beta(Q)\left[\int V^Q(q)dF(q)\right] \\ V_2^Q(x) = U(Q,x) + \beta(Q,q)\left[\int V^Q(q)dF(q)\right] \\ \int V^Q(q)dF(q) = \int_0^{q_{min}^Q} V_1^Q(q)dF(q) + \int_{q_{min}^Q}^{q_{max}^Q} V_2^Q(q)dF(q) + \int_{q_{max}^Q}^1 V_1^Q(q)dF(q) \end{cases}$$

It is easy to check that $\Pi(Q,0) > \Theta(Q,0)$ when β(Q,0)=β(Q) and with U(Q,0)<U(Q). Thus, q_{min}^Q will have a value in the range $[0,q_{max}^Q]$ if there exists a $q \in [0,q_{max}^Q]$ such as Π(Q,q) = Θ(Q,q). A sufficient condition for such a q to exist would be : $\Pi(Q,q_{max}^Q) < \Theta(Q,q_{max}^Q)$.Note that for $x = q_{max}^Q$,no collaboration is feasible, $V_1^Q(q)$becomes independent of q and$V_1^Q = \frac{U(Q)}{1-\beta(Q)}$. Thus :

$$\Pi\left(Q,q_{max}^Q\right) < \Theta\left(Q,q_{max}^Q\right)$$

$$\Leftrightarrow U(Q) + \beta(Q)\left[\int_0^1 V_1^Q(q)dF(q)\right] < U\left(Q,q_{max}^Q\right) + \beta(Q,q_{max}^Q)\left[\int_0^1 V_1^Q(q)dF(q)\right]$$

$$\Leftrightarrow U(Q) < U\left(Q,q_{max}^Q\right)(1-\beta(Q)) + \beta(Q,q_{max}^Q)U(Q)$$

$$\Leftrightarrow \frac{U(Q)}{1-\beta(Q)} < \frac{U\left(Q,q_{max}^Q\right)}{1-\beta(Q,q_{max}^Q)}$$

When this inequality is true, q_{min}^Q exists and is lower than q_{max}^Q.

Annex : Empirical results

Table 1: Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
hh	1566	3.210728	3.428919	0	29
gg	1566	7.43742	8.611137	0	83
nb_coauthors	1566	4.510217	7.447595	0	60
h	1566	3.251596	3.991914	0	39
g	1566	6.02235	8.130808	0	84
CL_index	1566	33.77775	96.90919	0	481.349
LLG_inde	1566	13.72596	29.03717	0	76.6924
age	1566	46.69796	10.20819	28	68
Prof_experience	1566	21.92593	9.360218	7	46
cohorte64_68	1566	.0127714	.1123225	0	1
cohorte69_73	1566	.0606641	.2387894	0	1
cohorte74_78	1566	.1264368	.3324471	0	1
cohorte79_83	1566	.1392082	.3462742	0	1
cohorte84_88	1566	.1085568	.3111818	0	1
cohorte89_93	1566	.1890166	.3916469	0	1
cohorte94_98	1566	.2273308	.4192419	0	1
cohorte99_04	1566	.1360153	.3429143	0	1
Female	1566	.2828863	.4505455	0	1
Male	1566	.7171137	.4505455	0	1
Full Prof.	1566	.3473819	.4762904	0	1
Ass. Prof.	1566	.6526181	.4762904	0	1

never_published	1566	.2247765	.4175684	0	1
cowrite_dr	1566	.1487867	.3559918	0	1
nb_papers	1566	8.360792	14.17236	0	157
nb_papers_Review	1566	6.182631	9.071662	0	94
nb_papers_other	1566	2.178161	6.367514	0	91
never_published_Rev	1566	.0197957	.139342	0	1
nb_papers_Rev_cnrs	1566	5.376756	8.205418	0	85
share_Rev_cnrs	1183	.8520602	.2479001	0	2
share_cnrs1	1133	.0679517	.1574878	0	1
share _cnrs2	1133	.1335283	.1982947	0	1
share _cnrs3	1133	.4583329	.3324206	0	1
share _cnrs4	1133	.3349856	.3631996	0	1
share _cnrs5	1133	.0052015	.0360094	0	.6
share_papers_fr	1183	.625243	.3329255	0	1
share_papers_gb	1183	.3474892	.3250533	0	1
share_papers_other	1183	.0272677	.1147627	0	1
working_alone_only	1566	.1232439	.328822	0	1
never_working_alone	1566	.1736909	.3789645	0	1
Herfindahl_JEL_code	1566	.3073441	.2636275	0	1
max_kwa	1566	0	0	0	0
max_kwb	1566	.0102171	.1005942	0	1
max_kwc	1566	.0006386	.0252699	0	1
max_kwd	1566	.0070243	.0835427	0	1
max_kwe	1566	.0102171	.1005942	0	1
max_kwf	1566	.0102171	.1005942	0	1
max_kwg	1566	.0019157	.0437408	0	1
max_kwh	1566	.0025543	.0504914	0	1

max_kwi	1566	.0025543	.0504914	0	1
max_kwj	1566	.0038314	.0617995	0	1
max_kwk	1566	0	0	0	0
max_kwl	1566	.0031928	.056433	0	1
max_kwm	1566	.0012771	.0357257	0	1
max_kwn	1566	0	0	0	0
max_kwo	1566	.0057471	.0756158	0	1
max_kwp	1566	.0031928	.056433	0	1
max_kwq	1566	.0025543	.0504914	0	1
max_kwr	1566	.0051086	.0713141	0	1
max_kwt	1566	0	0	0	0
max_kwy	1566	0	0	0	0
max_kwz	1566	.00447	.0667297	0	1
best_alone_C1	1566	.059387	.236423	0	1
best_alone_C2	1566	.0951469	.2935114	0	1
best_alone_C3	1566	.2662835	.4421555	0	1
best_alone_C4	1566	.1091954	.3119838	0	1
best_alone_C5	1566	.0006386	.0252699	0	1
best_alone_Econlit	1566	.0440613	.2052969	0	1
best_alone_misc	1566	.0268199	.1616085	0	1

Table 2: Correlations between individual research productivity indicators

	h	g	CL_index	LLG_index
h	1.0000			
g	0.9503*	1.0000		
CL_index	0.6636*	0.6372*	1.0000	
LLG_index	0.7302*	0.6838*	0.8970*	1.0000

* All correlations are significant at the 5% level. N=1566

Table 3: Descriptive Statistics by gender

gender	variable	mean	sd	min	max
Female	hh	2.505643	2.50553	0	12
	gg	5.744921	6.020361	0	35
	h	2.311512	2.686067	0	22
	g	4.383747	5.462493	0	46
	LLG_index	10.76408	16.95661	0	135.3494
	CL_index	14.62185	24.15478	0	175.3353
	Ass. PROF.	.8284424	.3774217	0	1
	Full PROF.	.1715576	.3774217	0	1
	nb_papers	5.187359	7.561131	0	72
	nb_coauthors	3.255079	4.722665	0	37
	nb_papers_Rev_cnrs	3.498871	5.056412	0	52
	age	42.15576	8.976371	29	66
Male	hh	3.488869	3.69499	0	29
	gg	8.105076	9.358078	0	83
	h	3.62244	4.34722	0	39
	g	6.668744	8.88703	0	84
	LLG_index	21.60745	40.15033	0	419.7203
	CL_index	41.33436	112.5501	0	1481.349
	Ass. PROF.	.5832591	.4932388	0	1
	Full PROF.	.4167409	.4932388	0	1
	nb_papers	9.612645	15.87744	0	157
	nb_coauthors	5.005343	8.22879	0	60
	nb_papers_Rev_cnrs	6.117542	9.049797	0	85
	age	48.48976	10.11188	28	68

Total	hh	3.210728	3.428919	0	29
	gg	7.43742	8.611137	0	83
	h	3.251596	3.991914	0	39
	g	6.02235	8.130808	0	84
	LLG_index	18.54001	35.5078	0	419.7203
	CL_index	33.77775	96.90919	0	1481.349
	Ass. PROF.	.6526181	.4762904	0	1
	Full. PROF.	.3473819	.4762904	0	1
	nb_papers	8.360792	14.17236	0	157
	nb_coauthors	4.510217	7.447595	0	60
	nb_papers_Rev_cnrs	5.376756	8.205418	0	85
	age	46.69796	10.20819	28	68

Table 4: Descriptive Statistics by academic position

academic_position	variable	mean	sd	min	max
ASS. PROF	hh	2.246575	2.34403	0	19
	gg	5.039139	5.545329	0	46
	h	2.016634	2.28411	0	25
	g	3.681018	4.688119	0	45
	LLG_index	9.572453	16.04467	0	159.139
	CL_index	11.69965	23.10019	0	244.0104
	nb_papers	4.466732	7.084793	0	87
	nb_coauthors	2.613503	4.187142	0	52
	nb_papers_Rev_cnrs	3.047945	4.380825	0	41
FULL PROF	hh	5.022059	4.303353	0	29
	gg	11.94301	11.16833	0	83
	h	5.571691	5.277773	0	39
	g	10.42096	10.93212	0	84

variable	mean	sd	min	max
LLG_index	35.38715	52.09662	0	419.7203
CL_index	75.25537	153.0455	0	1481.349
nb_papers	15.67647	20.0588	0	157
nb_coauteurs	8.073529	10.36365	0	60
nb_papers_Rev_cnrs	9.751838	11.33936	0	85

Table 5: Individual research productivity indicators per year and by cohorts

cohorte	variable	mean	sd	min	max
cohorte64-68	hh_peryear	.0923135	.0817156	0	.2608696
	gg_peryear	.2161584	.1898998	0	.6222222
	h_peryear	.1251731	.0997574	0	.3913043
	g_peryear	.2227694	.207504	0	.8260869
	LLG_index_peryear	.4454247	.5097879	0	2.037797
	CL_index_peryear	.8343734	1.033474	0	4.509589
	nb_papers_peryear	.2942663	.3880721	0	1.673913
	nb_coauthors_peryear	.1229395	.2368148	0	.9130435
	share_gb	.1710219	.1875821	0	.5555556
	share_fr	.7564798	.1953899	.3333333	1
	never_published	.2	.4103913	0	1
	female	.1	.3077935	0	1
cohorte69-73	hh_peryear	.0929661	.1111759	0	.5641026
	gg_peryear	.2210608	.2746659	0	1.589744
	h_peryear	.1056686	.1228009	0	.7435898
	g_peryear	.196483	.2491615	0	1.641026
	LLG_index_peryear	.4949158	.8717781	0	6.357573
	CL_index_peryear	1.135906	2.575137	0	19.68135
	nb_papers_peryear	.2589028	.399317	0	2.051282

	nb_coauthors_peryear	.1008967	.1949082	0	1.131579
	share_gb	.2084839	.2715722	0	1
	share_fr	.7277525	.3117807	0	1
	never_published	.2631579	.4426835	0	1
	female	.1052632	.3085203	0	1
cohorte74-78	hh_peryear	.0809532	.1051675	0	.71875
	gg_peryear	.2075475	.3257318	0	2.441176
	h_peryear	.0951696	.1210431	0	.7142857
	g_peryear	.1746357	.2553125	0	1.84375
	LLG_index_peryear	.4144785	.8375098	0	5.029756
	CL_index_peryear	.9674282	2.448345	0	18.8523
	nb_papers_peryear	.2262017	.444681	0	2.485714
	nb_coauthors_peryear	.0984183	.1981759	0	1.21875
	share_gb	.2348792	.2979328	0	1
	share_fr	.7313249	.3142713	0	1
	never_published	.2676768	.4438704	0	1
	female	.1111111	.3150663	0	1
cohorte79-83	hh_peryear	.0871213	.1201428	0	.6666667
	gg_peryear	.2099158	.2938712	0	1.666667
	h_peryear	.0912466	.1464147	0	.7407407
	g_peryear	.174917	.2953089	0	1.740741
	LLG_index_peryear	.434929	.9996627	0	8.562396
	CL_index_peryear	.8879347	2.604961	0	22.54853
	nb_papers_peryear	.2075978	.4104167	0	2.964286
	nb_coauthors_peryear	.1142663	.2397139	0	1.555556
	share_gb	.2380916	.2938016	0	1
	share_fr	.7153819	.32319	0	1
	never_published	.3899083	.4888517	0	1

	female	.2155963	.4121819	0	1
cohorte84-88	hh_peryear	.161737	.2078012	0	1.208333
	gg_peryear	.3747277	.4848179	0	2.772727
	h_peryear	.157683	.2389585	0	1.625
	g_peryear	.2884432	.4640665	0	3.5
	LLG_index_peryear	1.020523	2.418725	0	18.24871
	CL_index_peryear	2.574519	8.349001	0	61.72287
	nb_papers_peryear	.4309983	.882481	0	6.304348
	nb_coauthors_peryear	.2328522	.4375852	0	2.5
	share_gb	.3418605	.3280554	0	1
	share_fr	.6336796	.3337274	0	1
	never_published	.2882353	.4542793	0	1
	female	.2058824	.4055394	0	1
cohorte89-93	hh_peryear	.1753634	.1747176	0	.9473684
	gg_peryear	.3967337	.42326	0	2.190476
	h_peryear	.182196	.2249889	0	1.5
	g_peryear	.3379134	.4548101	0	3.555556
	LLG_index_peryear	1.202878	2.344683	0	20.15555
	CL_index_peryear	2.094658	5.73278	0	69.01778
	nb_papers_peryear	.5226029	.8686731	0	8.263158
	nb_coauthors_peryear	.2969896	.4593647	0	2.894737
	share_gb	.3594232	.31641	0	1
	share_fr	.6215	.3266946	0	1
	never_published	.2297297	.4213714	0	1
	female	.347973	.477134	0	1
cohorte94-98	hh_peryear	.2321876	.1929647	0	1.307692
	gg_peryear	.5267201	.4591398	0	2.923077

	h_peryear	.2192839	.1945358	0	1.3125
	g_peryear	.4058047	.4235059	0	3.066667
	LLG_index_peryear	1.320411	1.960414	0	18.70189
	CL_index_peryear	1.872301	3.235122	0	34.11071
	nb_papers_peryear	.5752364	.7581522	0	5.538462
	nb_coauthors_peryear	.3379599	.4525615	0	3.466667
	share_gb	.4117107	.329633	0	1
	share_fr	.5704174	.3349058	0	1
	never_published	.1404494	.3479417	0	1
	female	.3960674	.4897671	0	1
cohorte99-04	hh_peryear	.358184	.2572319	0	1.6
	gg_peryear	.7796462	.6420484	0	3.9
	h_peryear	.3126832	.2864607	0	2.1
	g_peryear	.5729656	.605065	0	5
	LLG_index_peryear	1.905336	2.112206	0	12.06301
	CL_index_peryear	2.099203	2.878418	0	18.47889
	nb_papers_peryear	.7562915	.8380927	0	6.4
	nb_coauthors_peryear	.4621283	.5000604	0	2.9
	share_gb	.4511526	.3277379	0	1
	share_fr	.5293931	.3284885	0	1
	never_published	.084507	.278802	0	1
	female	.3896714	.4888245	0	1
Total	hh_peryear	.1813883	.2000601	0	1.6
	gg_peryear	.4130868	.477887	0	3.9
	h_peryear	.1766795	.2154634	0	2.1
	g_peryear	.3265625	.4407232	0	5
	LLG_index_peryear	1.046137	1.915127	0	20.15555
	CL_index_peryear	1.712053	4.433391	0	69.01778

nb_papers_peryear	.4561684	.738855	0	8.263158
nb_coauthors_peryear	.2571403	.4146959	0	3.466667
share_gb	.3474892	.3250533	0	1
share_fr	.625243	.3329255	0	1
never_published	.2247765	.4175684	0	1

Table 6: First-Stage Estimations: Results for the individual research productivity variables

	(1)		(2)		(3)		(4)	
Dependant variable:	H		G		CL_index		LLG_index	
Model	ZINB		ZINB		Heckman Selection		Heckman Selection	
	IRR	P>|z|	IRR	P>|z|	Coeff.	P>|z|	Coeff.	P>|z|
Age	.988491*	0.056	.9926215	0.263	-.2711126	0.573	-.0020783	0.988
Female	.8630865**	0.014	.9326821	0.300	-9.133664***	0.005	-2.749063***	0.010
cohort64-68	.639886*	0.079	.5788128*	0.056	11.36906	0.523	-16.63519***	0.002
cohort69-73	.5865155***	0.010	.5446831***	0.006	29.23648*	0.103	-6.493137	0.218
cohort74-78	.5313103***	0.000	.5011169***	0.006	22.62574	0.115	-8.24296**	0.033
cohort79-83	.5673286***	0.001	.6145378***	0.008	29.77376**	0.030	-1.641273	0.676
cohort84-88	.6655879***	0.005	.6295683***	0.002	62.13896***	0.004	8.849404	0.110
cohort89-93	.7785253**	0.015	.7785826**	0.034	31.64009***	0.003	5.458532**	0.039
cohort94-98	.8393262**	0.035	.830293**	0.048	18.24714***	0.001	3.227771*	0.057
Working_alone_only	.3966277***	0.000	.4517704***	0.000	-41.56424***	0.000	-16.34336***	0.000
Cowrite_dr	1.417472***	0.000	1.427765***	0.000	9.533981	0.178	2.903077*	0.094
Best_alone_CNRS1	4.871913***	0.000	4.244295***	0.000	211.1036***	0.000	66.26043***	0.000
Best_alone_CNRS2	3.446526***	0.000	3.04466***	0.000	42.6116***	0.000	22.48153***	0.000
Best_alone_CNRS3	2.274868***	0.000	1.823829***	0.000	15.86695***	0.000	9.011317***	0.000
Best_alone_CNRS4	1.855814***	0.000	1.515237***	0.000	3.246687	0.369	1.191356	0.319
Best_alone_CNRS5	2.071003***	0.000	1.870775***	0.000	8.04704	0.191	3.649929**	0.046
Best_alone_Econlit_no_CNRS	2.027623***	0.000	1.915145***	0.000	9.57467	0.353	1.144091	0.614
Best_alone_Miscellaneous	2.759744***	0.000	2.396928***	0.000	.4210625	0.931	2.107959	0.208
Constant	.2143665***	0.000	.3815675***	0.000	-2.720161	0.869	-9.986753**	0.045

	Inflate : logit model				Selection model: nb_papers = 0			
	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|
Age	.2174416***	0.000	.1578476***	0.000	-.0646431***	0.000	-.0638421***	0.000
female	2.330623***	0.000	1.795803***	0.000	-.235634***	0.010	-.23097***	0.010
Working_alone_only	-2.034985	0.165	-.8145498	0.126	9.545215***	0.000	9.748365***	0.000
cohort64-68	.5527885	0.923	-.3973625	0.877	-1.085461**	0.033	-1.269771**	0.021
cohort69-73	14.68225***	0.007	1.028852	0.520	-.7991227**	0.015	-.8572475***	0.009
cohort74-78	15.23061***	0.004	1.682535	0.271	-.6080159**	0.018	-.6350715**	0.013
cohort79-83	17.04227***	0.001	3.065615**	0.035	-.5913037***	0.008	-.6033947***	0.007
cohort84-88	16.48781***	0.001	2.505392*	0.083	-.5503679**	0.012	-.5573013***	0.010
cohort89-93	16.68975***	0.001	2.753779**	0.045	-.2803182	0.106	-.2938074*	0.086
cohort94-98					-.1119017	0.490	-.1161586	0.466
PR_CE					2.724658***	0.000	2.961577***	0.000
PR_2C					1.832338***	0.000	1.873646***	0.000
PR_1C					1.236857***	0.000	1.225056***	0.000
MCF_HC					.5129406***	0.000	.5031609***	0.000
constant	-30.67345***	0.000	-12.73915***	0.000	3.595324***	0.000	3.549177***	0.000
lnalpha	-1.06703***	0.000	-.4579795***	0.000				
athrho					-.327532***	0.000	-.4281149***	0.000
N	1566		1566		1566		1566	
Log Likelihood	-3256.107		-4151.489		-7746.526		-6189.231	
Vuong Test (uncorrected) Vuong Test (AIC corrected)	3.58***	0.002	5.45***	0.000				
Likelihood-ratio test of alpha=0	589.77***	0.000	2741.61***	0.000				
Wald test of indep. eqns. (rho = 0)					48.44***	0.000	30.13***	0.000

The IRR value is the Incidence Rate Ratio of variable I and it is calculated as e^{β_i}; so if regressor i is increased by 1% the dependant variable will be increased by (1-IRR) %.P-values are reported in the P>|z| column (robust standard errors are calculated by bootstrap). The null hypothesis is rejected at the 1% level (***), 5 % (**) and 10% (*).lnalpha indicates the overdispersion parameter of the negative binomial distribution. The offset variable is the professional experience variable in each model. In the Heckman selection model, the athrho variable is the estimate of the inverse hyperbolic tangent of rho (□): athrho=0.5*ln((1+□)/(1-□)) where □ is the correlation between the residuals of the two equations. The Wald test of independent equations is the likelihood-ratio test of Ho: □ = 0 and it is computationally the comparison of the joint likelihood of an independent probit model for the selection equation and a regression model on research productivity index data against the Heckman model likelihood.

Table 7: Determinants of co-authorship: 2SRI Estimation Results

Dependant variable:	HH				GG				NB_COAUTHORS			
	(1)		(2)		(3)		(4)		(5)		(6)	
Model	ZIP		ZIP		ZINB		ZINB		ZINB		ZINB	
	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|
H	1.039757***	0.000	1.042237***	0.000					1.004025	0.688		
G					1.020522***	0.000	1.021633***	0.000			.9983641	0,727
Standardized-Residual	1.126451***	0.000	1.109018***	0.000	1.133134***	0.000	1.102424***	0.000	1.167272***	0.000	1.1922***	0.000
Age	.9852736***	0.000	.9854045***	0.000	.9906243**	0.023	.9888942***	0.007	.975415***	0.000	.9729548***	0.000
Female	.9241381**	0.038	.9306992*	0.059	.9840804	0.694	.985529	0.724	.9218153*	0.099	.9233338	0.113
Cowrite_dr	1.261285***	0.000	1.26846***	0.000	1.238887***	0.000	1.255716***	0.000	1.270383***	0.000	1.328744***	0.000
PR_CE	1.319049***	0.001	1.204146**	0.049	1.426929***	0.000	1.271924***	0.010	1.4579***	0.001	1.568339***	0.000
PR_2C	1.123684***	0.001	1.132486***	0.005	1.124639**	0.017	1.129122**	0.014	1.195797***	0.002	1.205116***	0.002
PR_1C	1.201338***	0.005	1.140478**	0.031	1.17343***	0.010	1.126719*	0.052	1.333257***	0.000	1.330581***	0.000
MCF_HC	.8917328	0.197	.8833541	0.188	.890631	0.155	.8942705	0.172	.9197727	0.454	.8825842	0.262
Working_alone_only	.531641***	0.000	.5334222***	0.000								
cohort64-68	.3690495***	0.000	.4155215***	0.000	.3122876***	0.000	.3633593***	0.000	.3394253***	0.000	.4128732***	0.001
cohort69-73	.4246863	0.000	.4468632***	0.000	.3755773***	0.000	.4010743***	0.000	.3581325***	0.000	.3702197***	0.000
cohort74-78	.3903951***	0.000	.4242865***	0.000	.4108896***	0.000	.4422769***	0.000	.3807995***	0.000	.3983065***	0.000
cohort79-83	.4608758***	0.000	.475241***	0.000	.44187***	0.000	.4694847***	0.000	.5035105***	0.000	.5392617***	0.000
cohort84-88	.5589912***	0.000	.5661185***	0.000	.5598416***	0.000	.5795794***	0.000	.6137121***	0.000	.6607299***	0.000
cohort89-93	.5962211***	0.000	.6119466***	0.000	.605443***	0.000	.6270194***	0.000	.7133021***	0.000	.7490796***	0.001
cohort94-98	.7079183***	0.000	.7329668***	0.000	.7258089***	0.000	.7541094***	0.000	.7670176***	0.000	.7864654***	0.000
nb_papers_Misc	1.006274***	0.003	1.005524***	0.006	1.00685**	0.023	1.007065**	0.019	1.016252***	0.000	1.021924***	0.000
nb_papers_EconLit_no_CNRS	1.004665	0.527	1.001992	0.774	1.009589	0.290	1.00799	0.370	1.03505***	0.001	1.03261***	0.000
nb_papers_CNRS1	.9893791**	0.027	.9897854**	0.040	.987247*	0.052	.9847426**	0.018	1.011172	0.118	1.005861	0.440
nb_papers_CNRS2	1.01123	0.217	1.010868	0.264	1.023509**	0.012	1.028064***	0.003	1.01944	0.079	1.024409**	0.034
nb_papers_CNRS3	1.006837	0.136	1.006638	0.146	1.013484***	0.006	1.013384***	0.007	1.05563***	0.000	1.060079***	0.000
nb_papers_CNRS4	1.00393	0.507	1.006825	0.214	1.002371	0.678	1.00501	0.395	1.053119***	0.000	1.050871***	0.000
nb_papers_CNRS5	.9734185	0.607	.9681081	0.565	1.022522	0.718	1.0128	0.836	.9457282	0.459	1.007413	0.924
Herfindahl_JEL_CODE	.8011764***	0.009	.8183969**	0.018	.7890423***	0.005	.7870318***	0.005	.4493046***	0.000	.4381293***	0.000
Share_gb	1.308479***	0.000	1.323982***	0.000	1.531048***	0.000	1.489122***	0.000	1.747907***	0.000	1.881927***	0.000
Share_other	1.09702	0.550	1.106846	0.509	1.108722	0.534	1.072947	0.673	.8145187	0.412	.8954796	0.663
PhD defended at (Network effect I) :												
Univ. of Toulouse 1	1.067021	0.413	1.090307	0.319	1.008725	0.917	1.008725	0.917	1.231663*	0.067	1.212608*	0.062
Other French research institution	.9622981	0.345	1.024923	0.610	.9336406	0.107	.9336406	0.107	.9717338	0.645	.9738609	0.621
Univ. of Paris 10	.9767195	0.670	.978346	0.712	.9826961	0.798	.9826961	0.798	.8433606*	0.064	.8400375**	0.049
Univ. of Aix-Marseille	1.072036	0.222	1.057769	0.464	1.089994	0.203	1.089994	0.203	.891283	0.260	.8785316	0.143
Univ. of Strasbourg	1.105696	0.227	1.125961	0.283	1.076385	0.417	1.076385	0.417	.7756882	0.141	.876032	0.255
Univ. of Paris 9	.7674918**	0.023	.7374286***	0.007	.8468169*	0.095	.8468169*	0.095	.9912081	0.943	1.01117	0.928
Grande Ecole	1.182339**	0.014	1.155048*	0.056	1.309115***	0.001	1.309115***	0.001	1.003747	0.971	1.064606	0.536
Other Univ. In Paris	1.098862*	0.091	1.052457	0.359	1.108811	0.136	1.108811	0.136	.9464552	0.535	1.008636	0.924
European country	1.395722***	0.007	1.439115***	0.002	1.560409***	0.002	1.560409	0.002	1.38816*	0.060	1.268299	0.191
US	.890038	0.523	1.002826	0.985	.9419018	0.760	.9419018	0.760	1.52725*	0.056	1.47037*	0.099
Universities (Network effect II)	NO		Yes		NO		Yes		Yes		NO	

Inflate : logit model	(1)		(2)		(3)		(4)		(5)		(6)	
	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|
H	-3.414301***	0.001	-3.37412***	0.010					-7.548156***	0.000		
G					-.7498065***	0.000	-.7498065***	0.000			-5.219176***	0.000
Standardized-Residual	1.736803*	0.107	1.605957	0.186	1.465291**	0.011	1.465291**	0.011	14.75788***	0.000	22.05012***	0.000
Age	-.0435219	0.454	-.0350149	0.544	.0178888	0.527	.0178888	0.527	-.1305535*	0.063	-.5140615***	0.003
Female	-.1805496	0.753	-.1502756	0.808	.0666068	0.818	.0666068	0.819	-38.75202	0.952	-37.45614	0.936
PR_CE	.9114115	0.784	.913769	0.805	-.4992077	0.642	-.4992077	0.642	1.603895	0.168	1.160897	0.383
PR_2C	-.4850895	0.506	-.5655367	0.484	.2461645	0.516	.2461645	0.516	.6723288	0.380	.5102187	0.627
PR_1C	-.4845668	0.544	-.5105686	0.553	-.6838853*	0.101	-.6838853*	0.101	1.402785**	0.047	1.875419*	0.088
MCF_HC	-1.272412	0.284	-1.119334	0.326	.3343665	0.421	-.3343665	0.421	.3779083	0.566	-.0398384	0.967
cohort64-68	6.98064**	0.023	6.837184**	0.035	2.612548	0.167	2.612548	0.167	32.54643***	0.000	75.41312	0.973
cohort69-73	4.503021**	0.024	4.380308**	0.033	1.906422*	0.057	1.906422*	0.057	19.76823***	0.001	63.7169	0.978
cohort74-78	2.326547*	0.099	2.382214	0.125	1.48731**	0.050	1.48731**	0.050	15.54651***	0.002	57.10527	0.980
cohort79-83	1.490585	0.258	1.364695	0.322	.5237884	0.472	.5237884	0.472	14.23103***	0.002	55.29417	0.981
cohort84-88	1.874462	0.184	1.70459	0.301	.7831614	0.230	.7831614	0.230	15.30598***	0.001	54.52447	0.981
cohort89-93	1.40844	0.178	1.415473	0.229	.4452705	0.436	.4452705	0.436	12.68754***	0.003	45.61329	0.984
cohort94-98	.6074362	0.475	.6644003	0.519	.2522305	0.607	.2522305	0.607	9.68772***	0.004	19.89594	0.993
constant	2.420922	0.390	1.853168	0.525	-.6780755	0.631	-.6780755	0.631	4.485352	0.215	-10.7505	0.993
lnalpha					-1.952367***	0.000	-1.952367***	0.000	-1.901721***	0.000	-1.568005***	0.000
N	1183		1183		1183		1183		1183		1183	
Log Likelihood	-2262.49		-2213.927		-3272.022		-3836.815		-2601.264		-2680.350	
Vuong Test (unconstraint)	4.73***	0.000	4.38***	0.000	8.28***	0.000	8.10***	0.000	6.56***	0.000	6.55***	0.000
Likelihood-ratio test of alpha=0					575.91 ***	0.000	352.01***	0.000	370.94***	0.000	553.21***	0.000
Exogeneity test (Wald test)	32.94μμμ	0.000	20.93μμμ	0.000	29.61μμμ	0.000	15.94μμμ	0.000	42.85μμμ	0.000	45.05μμμ	0.000

The IRR value is the Incidence Rate Ratio of variable i ant it is calculated as e^{β_i}; so if regressor i is increased by 1% the dependant variable will be increased by (1-IRR) %. P-values are reported in the P>|z| column (robust standard errors are calculated by bootstrap). The null hypothesis is rejected at the 1% level (***), 5 % (**) and 10% (*).lnalpha indicates the overdispersion parameter of the negative binomial distribution. The null hypothesis of exogeneity is rejected at the 1% level (μμμ), 5% level (μμ) and 10% level(μ). All parent models include a constant parameter which is not reported in the table. The offset variable is the Experience variable.

Table 8: Robustness checks: Determinants of co-authorship

Dependant variable:	HH				GG				NB_COAUTHORS			
	(1)		(2)		**(3)**		**(4)**		**(5)**		**(6)**	
Model	ZINB		ZINB		ZINB		ZINB		ZINB		ZINB	
	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|
CL-index	1.0016***	0.000			1.001304***	0.000			1.002478***	0.000		
LLG-index			1.005915***	0.000			1.004632***	0.000			1.005347***	0.000
Standardized-Residual	1.003008	0.691	1.007984**	0.022	1.015563**	0.038	1.011484***	0.002	1.057704***	0.000	1.03337***	0.000
Age	.9772159	0.000	.9792032***	0.000	.991479*	0.078	.9928148	0.123	.9592894***	0.000	.9705186***	0.000
Female	.9013064**	0.036	.9091849**	0.044	.9703375	0.542	.9898528	0.827	.779914***	0.000	.8767449***	0.019
Cowrite_dr	1.366051	0.000	1.37816***	0.000	1.275365***	0.000	1.274232***	0.000	1.357346***	0.000	1.32112***	0.000
PR_CE	2.389134***	0.000	1.928204***	0.000	1.874848***	0.000	1.621036***	0.000	2.652294***	0.000	1.990655***	0.000
PR_2C	1.352706***	0.000	1.301035***	0.000	1.23523***	0.000	1.196042***	0.001	1.443428***	0.000	1.257602***	0.001
PR_1C	1.708428***	0.000	1.523754***	0.000	1.427667***	0.000	1.313868***	0.000	1.958431***	0.000	1.58221***	0.000
MCF_HC	.9944974	0.955	.9465338	0.569	.3269134***	0.000	.8734581	0.147	.9103242	0.472	.9019052	0.400
cohort64-68	.3022697***	0.000	.328389***	0.000	.3839133***	0.000	.3354482***	0.000	.5507499*	0.059	.5376615**	0.031
cohort69-73	.337953***	0.000	.3543386***	0.000	.4138367***	0.000	.3902346***	0.000	.4179275***	0.000	.4084654***	0.000
cohort74-78	.3453074***	0.000	.3533337***	0.000	.4932791***	0.000	.4081982***	0.000	.5364283***	0.001	.4928987***	0.000
cohort79-83	.4105089***	0.000	.4148241***	0.000	.5715696***	0.000	.4786778***	0.000	.7452103*	0.088	.6635999***	0.008
cohort84-88	.5303776***	0.000	.5384428***	0.000	.6581819***	0.000	.5646688***	0.000	.7358037**	0.038	.7110265**	0.011
cohort89-93	.6078139***	0.000	.5976638***	0.000	.7476915****	0.000	.6373786***	0.000	.9940561	0.956	.9207603	0.395
cohort94-98	.7065424***	0.000	.7030873***	0.000	.8242373***	0.000	.7241441***	0.000	.9793003	0.811	.9057413	0.198
PhD defended at (Network effect I) :												
Univ. of Toulouse 1	.8588143	0.119	.8906226	0.218	.8242373*	0.052	.8774745	0.175	.8375818	0.173	.8475816	0.161
Other French research institution	.8947545**	0.027	.9215736*	0.095	.8568114***	0.002	.880456***	0.009	.886165*	0.064	.9329208	0.248
Univ. of Paris 10	.9984095	0.984	1.000657	0.993	.9067076	0.223	.9152171	0.256	.7922757**	0.030	.78888**	0.017
Univ. of Aix-Marseille	.965878	0.666	.9764135	0.761	.9841681	0.844	.9910739	0.909	.8162309*	0.059	.8223728**	0.048
Univ. of Strasbourg	1.281718**	0.015	1.273284**	0.013	1.099083	0.388	1.10262	0.355	1.013742	0.921	.9879296	0.925
Univ. of Paris 9	.7912667**	0.040	.7929879**	0.036	.8063962*	0.063	.8264638*	0.088	.8250456	0.199	.8278389	0.166
Grande Ecole	1.292688***	0.006	1.318144***	0.002	1.319735***	0.000	1.358476***	0.001	1.106039	0.421	1.041512	0.727
Other Univ. In Paris	1.091465	0.266	1.100146	0.214	1.056966	0.506	1.069775	0.402	1.034177	0.761	1.010941	0.915
European country	1.889908***	0.000	1.987807***	0.000	1.813658***	0.001	1.876908***	0.000	1.513432*	0.076	1.584631**	0.027
US	1.083623	0.724	1.224009	0.352	.9617424	0.869	1.048616	0.834	1.22257	0.501	1.355687	0.256
Herfindahl_KW	.7764863***	0.006	.7905826***	0.010	.6078884***	0.000	.6587948***	0.000	.3408761***	0.000	.3240986***	0.000
Share_gb					1.903197***	0.000	1.89281***	0.000			1.934445***	0.000
Share_other					1.134149	0.509	1.179378	0.369			.8166748	0.454

Inflate : logit model	(1)		(2)		**(3)**		**(4)**		**(5)**		**(6)**	
	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|
CL-index	-.1713406***	0.000			-.0353674***	0.000			-.4324095***	0.000		
LLG-index			-.400623***	0.000			-.0787158***	0.000			-2.6447***	0.000
Standardized-Residual	.095492*	0.088	.0787085*	0.074	.3386137***	0.000	.1747955***	0.000	1.374328***	0.000	4.515943***	0.000
Age	.0508537***	0.010	.0537892***	0.005	.045454*	0.080	.0689766***	0.006	-.0121094	0.579	.0969078	0.245
Female	.3428557	0.140	.3750159*	0.103	.3049845	0.291	.4559854*	0.095	-1.76851***	0.000	-6.621425***	0.000
PR_CE	-1.04646	0.117	-1.134538	0.133	-1.756188*	0.060	-1.644496**	0.045	-.783238	0.339	-1.219885	0.692
PR_2C	-.7530014*	0.073	-.6183909	0.131	.117625	0.756	-.1741986	0.637	-.9368175**	0.025	.2196244	0.869
PR_1C	-1.321026***	0.000	-1.312619***	0.001	-1.066789**	0.012	-1.141776***	0.005	-.1840026	0.666	.4426868	0.693
MCF_HC	-.2544673	0.389	-.3414806	0.265	-.505504	0.240	-.5819776	0.158	-.0793486	0.818	.2285113	0.843
cohort64-68	3.007732**	0.011	2.885046**	0.018	3.174666**	0.025	2.068796	0.121	11.6493***	0.000	35.07727***	0.000
cohort69-73	1.794842**	0.021	1.768232**	0.022	2.707837***	0.006	1.965276**	0.036	8.646351***	0.000	29.04559***	0.000
cohort74-78	1.603819**	0.024	1.508179**	0.029	2.912361***	0.001	2.025999**	0.011	9.815295***	0.000	32.86846***	0.000
cohort79-83	1.430626**	0.031	1.281215**	0.042	2.37005***	0.004	1.523386**	0.049	10.26298***	0.000	33.82659***	0.000
cohort84-88	1.587691***	0.008	1.468928**	0.011	2.471345***	0.001	1.676028**	0.021	8.135476***	0.000	27.14741***	0.000
cohort89-93	1.173924**	0.029	1.118926**	0.037	1.816784***	0.009	1.274693*	0.052	8.024847***	0.000	29.14987***	0.000
cohort94-98	.5039553	0.337	.407361	0.421	1.213887**	0.043	.5977456	0.301	5.777882***	0.000	13.93092***	0.000
constant	-3.612173***	0.000	-3.586909***	0.000	-4.673654***	0.000	-5.488118***	0.000	-.2726039	0.754	-12.0514***	0.001
lnalpha	-1.936493***	0.000	-2.10015***	0.000	-1.357579***	0.000	-1.46246***	0.000	-.8564862***	0.000	-1.112601***	0.000
N	1566		1566		1183		1183		1566		1183	
Log Likelihood	-3128.945		-3094.878		-3469.983		-3457.168		-3130.962		-2739.573	
Vuong Test	8.14***	0.000	8.27***	0.000	6.91***	0.000	7.08***	0.000	14.38***	0.000	10.99***	0.000
Likelihood-ratio test of alpha=0	157.60***	0.000	115.50***	0.000	1204.56***	0.000	1070.03***	0.000	1480.87***	0.000	1162.63	0.000
Exogeneity test (Wald test)	2.92	0.232	$7.58^{\mu\mu}$	0.0226	$38.47^{\mu\mu\mu}$	0.000	$29.47^{\mu\mu\mu}$	0.000	$139.44^{\mu\mu\mu}$	0.000	$75.36^{\mu\mu\mu}$	0.000

The IRR value is the Incidence Rate Ratio of variable i ant it is calculated as e^{β_i}; so if regressor i is increased by 1% the dependant variable will be increased by (1-IRR) %. P-values are reported in the P>|z| column (robust standard errors are calculated by bootstrap). The null hypothesis is rejected at the 1% level (***), 5 % (**) and 10% (*).lnalpha indicates the overdispersion parameter of the negative binomial distribution. The null hypothesis of exogeneity is rejected at the 1% level (μμμ), 5% level (μμ) and 10% level(μ). All parent models include a constant parameter which is not reported in the table. The offset variable is the Experience variable.